

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 16:44:17 ; Search time 21 Seconds  
(without alignments)  
678.989 Million cell updates/sec

Title: US-10-088-872-2  
Perfect score: 1704  
Sequence: 1 MKKMPLFSKSHKNPAEIVKI.....FADEKNYLIKQIRDLKKTAP 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1704	100.0	337	3	US-09-190-965-1	Sequence 1, Appli
2	1704	100.0	337	4	US-09-470-253-1	Sequence 1, Appli
3	1376	80.8	341	3	US-09-190-965-3	Sequence 3, Appli
4	1376	80.8	341	4	US-09-470-253-3	Sequence 3, Appli
5	1109	65.1	339	3	US-09-190-965-4	Sequence 4, Appli
6	1109	65.1	339	4	US-09-470-253-4	Sequence 4, Appli
7	1063.5	62.4	377	3	US-09-190-965-5	Sequence 5, Appli
8	1063.5	62.4	377	4	US-09-470-253-5	Sequence 5, Appli
9	128.5	7.5	3878	4	US-09-914-259-11	Sequence 11, Appli
10	113.5	6.7	1279	4	US-09-724-517-2	Sequence 2, Appli
11	113.5	6.7	1279	4	US-09-641-807A-2	Sequence 2, Appli

12	113.5	6.7	1279	4	US-09-723-096-2	Sequence 2, Appli
13	113	6.6	2184	4	US-09-417-485D-6	Sequence 6, Appli
14	105	6.2	586	2	US-08-630-822A-70	Sequence 70, Appl
15	105	6.2	586	2	US-09-005-069-70	Sequence 70, Appl
16	105	6.2	586	4	US-09-171-156A-30	Sequence 30, Appl
17	105	6.2	586	4	US-09-004-730A-30	Sequence 30, Appl
18	105	6.2	586	4	US-08-981-799A-30	Sequence 30, Appl
19	103.5	6.1	245	4	US-09-399-913-4	Sequence 4, Appli
20	103.5	6.1	245	4	US-09-298-731-4	Sequence 4, Appli
21	103	6.0	387	4	US-09-328-352-5367	Sequence 5367, Ap
22	103	6.0	2662	4	US-09-595-684B-31	Sequence 31, Appl
23	102.5	6.0	975	4	US-09-914-259-19	Sequence 19, Appl
24	102.5	6.0	1098	3	US-08-923-992A-8	Sequence 8, Appli
25	102.5	6.0	1164	3	US-08-923-992A-10	Sequence 10, Appl
26	102.5	6.0	1388	4	US-09-572-191-2	Sequence 2, Appli
27	102.5	6.0	1388	4	US-09-723-262-2	Sequence 2, Appli
28	102.5	6.0	1388	4	US-09-723-219-2	Sequence 2, Appli
29	102	6.0	474	3	US-08-387-117-6	Sequence 6, Appli
30	102	6.0	1128	3	US-08-923-992A-6	Sequence 6, Appli
31	101	5.9	1147	3	US-08-470-260-5	Sequence 5, Appli
32	101	5.9	1147	3	US-08-471-491-5	Sequence 5, Appli
33	101	5.9	1147	3	US-08-466-662-5	Sequence 5, Appli
34	101	5.9	3289	2	US-08-477-451-2	Sequence 2, Appli
35	99.5	5.8	1164	3	US-08-923-992A-2	Sequence 2, Appli
36	99	5.8	323	4	US-09-134-001C-3133	Sequence 3133, Ap
37	98	5.8	1048	3	US-09-356-952-5	Sequence 5, Appli
38	97	5.7	2482	1	US-08-328-254-6	Sequence 6, Appli
39	97	5.7	3248	1	US-08-353-700-1	Sequence 1, Appli
40	97	5.7	3248	5	PCT-US95-16216-1	Sequence 1, Appli
41	96	5.6	1183	4	US-09-107-532A-6680	Sequence 6680, Ap
42	95.5	5.6	967	4	US-09-914-259-21	Sequence 21, Appl
43	95.5	5.6	1027	4	US-09-914-259-27	Sequence 27, Appl
44	95	5.6	564	4	US-09-198-452A-601	Sequence 601, App
45	95	5.6	956	4	US-09-914-259-17	Sequence 17, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-190-965-1

; Sequence 1, Application US/09190965

; Patent No. 6071721

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Guegler, Karl J.

; APPLICANT: Corley, Neil C.

; APPLICANT: Gorgone, Gina A.

; TITLE OF INVENTION: CALCIUM BINDING PROTEIN

; FILE REFERENCE: PF-0635 US

; CURRENT APPLICATION NUMBER: US/09/190,965

; CURRENT FILING DATE: 1998-11-13

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PERL Program

; SEQ ID NO 1

; LENGTH: 337

; TYPE: PRT



US-09-470-253-1

Query Match 100.0%; Score 1704; DB 4; Length 337;  
Best Local Similarity 100.0%; Pred. No. 1.6e-161;  
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60

Qy     61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120

Qy    121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180

Qy    181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240

Qy    241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQPK 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQPK 300

Qy    301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
        ||||||||||||||||||||||||||||||||||||||||||||
Db    301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
```

### RESULT 3

US-09-190-965-3

; Sequence 3, Application US/09190965  
; Patent No. 6071721  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Gorgone, Gina A.  
; TITLE OF INVENTION: CALCIUM BINDING PROTEIN  
; FILE REFERENCE: PF-0635 US  
; CURRENT APPLICATION NUMBER: US/09/190,965  
; CURRENT FILING DATE: 1998-11-13  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PERL Program  
; SEQ ID NO 3  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: Mus sp.  
; FEATURE: -  
; OTHER INFORMATION: g262934  
US-09-190-965-3

Query Match 80.8%; Score 1376; DB 3; Length 341;  
Best Local Similarity 80.7%; Pred. No. 7.3e-129;  
Matches 272; Conservative 32; Mismatches 29; Indels 4; Gaps 2;



Db 1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy 60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119  
 || ||||| |||||: |||| ||: ||||| ||||| ||||| |||||: |||||

Db 61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy 120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179  
 : ||||| |||||: |||| ||||| ||||| ||||| |||||: |||||

Db 121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy 180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239  
 ||||| ||||| |||||: ||||| ||||| ||||| ||||| ||||| |||||: |||||

Db 181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLDDR 240

Qy 240 HNFAMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQP 299  
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||: |||||

Db 241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVVFVANPNKTQPILDILLKNQT 300

Qy 300 KLIEFLSSSQKERTDDEQFADEKNYLIKQIRDLKKTAA 336  
 ||||| ||| : ||| |||| ||| |||: |||||: |||

Db 301 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRNLKRAA 337

# RESULT 5

US-09-190-965-4

; Sequence 4, Application US/09190965

; Patent No. 6071721

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Guegler, Karl J.

; APPLICANT: Corley, Neil C.

; APPLICANT: Gorgone, Gina A.

; TITLE OF INVENTION: CALCIUM BINDING PROTEIN

; FILE REFERENCE: PF-0635 US

; CURRENT APPLICATION NUMBER: US/09/190,965

; CURRENT FILING DATE: 1998-11-13

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PERL Program

; SEQ ID NO 4

; LENGTH: 339

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

; FEATURE: -

; OTHER INFORMATION: g1794137

US-09-190-965-4

Query Match 65.1%; Score 1109; DB 3; Length 339;

Best Local Similarity 65.0%; Pred. No. 2.8e-102;

Matches 217; Conservative 59; Mismatches 54; Indels 4; Gaps 3;

Qy 4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63  
 |||| ||| :| :|| ||: : ||| :| :||| :| :||| :| :|||

Db 1 MPLFGKSQKSPVELVKSLKEAINALEAGDRKVEKAQEDVSKNLVSIKNMLHCGSSDAEPPA 60

Qy 64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122  
 : |||||: |||||: ||||| ||||| ||||| ||||| ||||| |||||

Db 61 DYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTVEYICTK 120

Qy 123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180  
 | ||| |: ||| |: ||| | |||| |: | ||||: | ::| |: |||: |||||  
 Db 121 PEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180  
 Qy 181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLQSENYVTKRQSLKLLGELILDR 239  
 ||||: |||: |||||: | |: ||| || | : |: || | |||||: |||||: |||||: |||  
 Db 181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLNSENIVTRRQSLKLLGELLDR 240  
 Qy 240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFASPHKTQPIVEILLKNQP 299  
 ||| : ||: |||: |||||: ||: || | |||||: |||||: ||: ||: |||: |||  
 Db 241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVVFANPNKPKPILDILLRNQT 300  
 Qy 300 KLIEFLSSSFQKERTDDEQFADEKNYLIKQIRDLK 333  
 ||: |||: | : ||: |||| ||| |||||: |||  
 Db 301 KLVDFLTNTFHTDRSEDEQFNDEKAYLIKQIKELK 334

# RESULT 6

US-09-470-253-4

; Sequence 4, Application US/09470253  
 ; Patent No. 6365371  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Guegler, Karl J.  
 ; APPLICANT: Corley, Neil C.  
 ; APPLICANT: Gorgone, Gina A.  
 ; TITLE OF INVENTION: CALCIUM BINDING PROTEIN  
 ; FILE REFERENCE: PF-0635 US  
 ; CURRENT APPLICATION NUMBER: US/09/470,253  
 ; CURRENT FILING DATE: 1999-12-22  
 ; PRIOR APPLICATION NUMBER: 09/190,965  
 ; PRIOR FILING DATE: 1998-11-13  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 4  
 ; LENGTH: 339  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; FEATURE: -  
 ; OTHER INFORMATION: g1794137

US-09-470-253-4

Query Match 65.1%; Score 1109; DB 4; Length 339;  
 Best Local Similarity 65.0%; Pred. No. 2.8e-102;  
 Matches 217; Conservative 59; Mismatches 54; Indels 4; Gaps 3;

Qy 4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63  
 |||| || |: | |: | |: | | |: | |: | |: | |: | |: | |: | |: |  
 Db 1 MPLFGKSQKSPVELVKSLKEAINALEAGDRKVEKAQEDVSKNLVSIKNMLHGSSDAEPPA 60  
 Qy 64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTEYISAH 122  
 : ||||: ||||: | |: | | | ||||| | ||||: |||||: |||||: |||||  
 Db 61 DYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTEYICTK 120  
 Qy 123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180  
 | ||| |: ||| |: ||| | |||| |: | ||||: | ::| |: |||: |||||

Db 121 PEILFTLMAGYEDAHPEIALNSGTMRLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180

QY 181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLQSENYVTKRQSLKLLGELILDR 239  
 |||:||||:|||||:| |:|: ||| | : |:| | |||||:|||||:||||

Db 181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLNSENIVTRRQSLKLLGELLDR 240

QY 240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVSPHKTQPIVEILLKNQP 299  
 ||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

Db 241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVVFVANPNKPKPILDILLRNQT 300

QY 300 KLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLK 333  
 ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

Db 301 KLVDFLTNFHTDRSEDEQFNDEKAYLIKQIKELK 334

# RESULT 7

US-09-190-965-5

; Sequence 5, Application US/09190965  
 ; Patent No. 6071721  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Guegler, Karl J.  
 ; APPLICANT: Corley, Neil C.  
 ; APPLICANT: Gorgone, Gina A.  
 ; TITLE OF INVENTION: CALCIUM BINDING PROTEIN  
 ; FILE REFERENCE: PF-0635 US  
 ; CURRENT APPLICATION NUMBER: US/09/190,965  
 ; CURRENT FILING DATE: 1998-11-13  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 5  
 ; LENGTH: 377  
 ; TYPE: PRT  
 ; ORGANISM: Caenorhabditis elegans  
 ; FEATURE: -  
 ; OTHER INFORMATION: g1255838  
 US-09-190-965-5

Query Match 62.4%; Score 1063.5; DB 3; Length 377;  
 Best Local Similarity 60.5%; Pred. No. 1.1e-97;  
 Matches 211; Conservative 53; Mismatches 68; Indels 17; Gaps 3;

QY 4 MP-LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAM 49  
 || || ||||:|:|:| | |:| | ||:| || || | |||:|:|

Db 1 MPLLFGKSHKSPADVVKTLREVLTILDKLPPPKLDDKGNISDKKYDKALDEVSKNVAMI 60

QY 50 KEILCGTNEKEPPTE---AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR 106  
 | : | : ||:| |||||:|:|:| || | :|| ||| ||||:|

Db 61 KSFIYGNDSEPSSEHVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGGQIFNNLLR 120

QY 107 RQIGTRSPTEVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFR 166  
 |||||:|:|:| | || | |:| | ||| |:| | |||:|:|:|

Db 121 RQIGTRSPTEYLGARPEILIQLVQGYSPDIALTCGLMLRESIRHDHLAKIILYSDFVY 180

QY 167 DFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKR 226  
 || ||: |||:||||:|:|:| |||| |:|:|: |||| | : || |:|:|:|

Db 181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAI IAEFLDSNYDTFFAQYQNLNLSKNYVTRR 240



QY 227 QSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHK 286  
 |||||:|||| | |||| |::|| | |||| |::|||:|||||:|:  
 Db 241 QSLKLLGELLLDRHNFTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVVFVANPNK 300  
 QY 287 TQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKT 335  
 :|| :|| :|: ||:|||| | :||||| || | |||||::| :  
 Db 301 PKPISDILNRNREKLVEFLSEFHNDRTDDEQFNDEKAYLIKQIQEMKSS 349

RESULT 8

US-09-470-253-5  
 ; Sequence 5, Application US/09470253  
 ; Patent No. 6365371  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Guegler, Karl J.  
 ; APPLICANT: Corley, Neil C.  
 ; APPLICANT: Gorgone, Gina A.  
 ; TITLE OF INVENTION: CALCIUM BINDING PROTEIN  
 ; FILE REFERENCE: PF-0635 US  
 ; CURRENT APPLICATION NUMBER: US/09/470,253  
 ; CURRENT FILING DATE: 1999-12-22  
 ; PRIOR APPLICATION NUMBER: 09/190,965  
 ; PRIOR FILING DATE: 1998-11-13  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 5  
 ; LENGTH: 377  
 ; TYPE: PRT  
 ; ORGANISM: Caenorhabditis elegans  
 ; FEATURE: -  
 ; OTHER INFORMATION: g1255838  
 US-09-470-253-5

Query Match 62.4%; Score 1063.5; DB 4; Length 377;  
 Best Local Similarity 60.5%; Pred. No. 1.1e-97;  
 Matches 211; Conservative 53; Mismatches 68; Indels 17; Gaps 3;

QY 4 MP-LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAM 49  
 || || ||||:|::|| |:: | ||:| || || |::|:: :  
 Db 1 MPLLFGKSHKSPADVVKTLREVLTILDKLPKPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60  
 QY 50 KEILCGTNEKEPPTE---AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR 106  
 | : | : || :| |||||:|:: :| || | :|| |||| | ||||:|  
 Db 61 KSFIYGNDSEPSSEHVQVQAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNILR 120  
 QY 107 RQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFR 166  
 |||||:| | || |::|| | || |::||| ||: |||||:|:  
 Db 121 RQIGTRSPTVEYLGARPEILIQLVQGYSPDIALTCGLMLRESIRHDHLAKIILYSDVIFY 180  
 QY 167 DFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKR 226  
 || ||: |||:||||:|::| |||| :|::|: |||| | : || |::|::|  
 Db 181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIEFLDSNYDTFFAQYQNLLNSKNYVTRR 240  
 QY 227 QSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHK 286  
 |||||:|||| | |||| |::|| | |||| |::|||:|||||:|:

Db 241 QSLKLLGELLDDRHNFTMTKYISNPDNLRLMMELLRDKSRNIQYEFHVFKVVFVANPNK 300

Qy 287 TQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKT 335  
 :|| :|| :| :||| | :||| | | | | | :| :|

Db 301 PKPISDILNRNREKLVEFLSEFHNDRTDDEQFNDEKAYLIKQIQEMKSS 349

RESULT 9

US-09-914-259-11  
 ; Sequence 11, Application US/09914259  
 ; Patent No. 6495336  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Makowski, Lee  
 ; APPLICANT: Hyman, Paul  
 ; APPLICANT: Williams, Mark  
 ; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
 ; FILE REFERENCE: 8471-010-999  
 ; CURRENT APPLICATION NUMBER: US/09/914,259  
 ; CURRENT FILING DATE: 2000-11-21  
 ; NUMBER OF SEQ ID NOS: 180  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 11  
 ; LENGTH: 3878  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-914-259-11

Query Match 7.5%; Score 128.5; DB 4; Length 3878;  
 Best Local Similarity 20.1%; Pred. No. 0.0037;  
 Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;

Qy 18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77  
 : : |||| | | : | :| : : | : || |

Db 664 IEKLDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE----- 710

Qy 78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI 125  
 : : : ||| | : : : : || | | : : | : : :

Db 711 --ISKLDLQOSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 766

Qy 126 LFMLLLGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185  
 | : | | : : | : | |

Db 767 LEKQMKKE-----NDLQEKFAQLEAEN-SILKDEKK 797

Qy 186 TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237  
 | :| :| | : : | : : : | :| :| | :| :| :|

Db 798 TLEDMLKIHTPVSQEERLIFLDSIKSKSDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 857

Qy 238 DRHNFAIMTK-----YISKPENLKLMMNLLRD 264  
 | : | : | : | : | :| :|

Db 858 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 917

Qy 265 KSPNIQFEA--FHVFKVVFVSPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320  
 : : : | | ||| :| :| :| :| :| :| :| :| :| :|

Db 918 NPTTVKMKSSVFDKTFVA---ETLEMGEVVEKDTTELMKLEVTREKLELSQRLSDL 974

Qy 321 -----EKNYLIKQIRDLKK 334  
 | :| : : : | | :



QY 252 PENLKLMMNLLRDKSPNIQFEAFHVFKVVFVSPHKTQPIVEILLKNQPKLIEFLSSF--- 308  
 ||::|| |:: | || :||| : | :: |:: ||  
 Db 1145 PESMKLSG---REREMDSS-----ASSLRTQPNPQKLWEDIPELPPIHSSLAPP 1190  
 QY 309 -----QKERTDDEQFADEKKNYLIKQIR 330  
 | ||| || : | ||:  
 Db 1191 SGHMLGNENKTETDDNQFTKSHSRLSSQIQ 1220

RESULT 11

US-09-641-807A-2  
 ; Sequence 2, Application US/09641807A  
 ; Patent No. 6440731  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Beraud, Christophe  
 ; APPLICANT: Freedman, Richard  
 ; TITLE OF INVENTION: No. 6440731el motor proteins and methods for  
 ; TITLE OF INVENTION: their use  
 ; FILE REFERENCE: 1031  
 ; CURRENT APPLICATION NUMBER: US/09/641,807A  
 ; CURRENT FILING DATE: 2000-08-17  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 1279  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: (409)...(446)  
 ; OTHER INFORMATION: Xaa = any amino acid  
 US-09-641-807A-2

Query Match 6.7%; Score 113.5; DB 4; Length 1279;  
 Best Local Similarity 19.3%; Pred. No. 0.024;  
 Matches 87; Conservative 61; Mismatches 137; Indels 165; Gaps 14;

QY 23 DNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPTEAVAQLAQELYSSGLLVTL 82  
 |:| |::| | : |::| :|| :|| | |  
 Db 794 DHLQKLDEQKKWLDEEVEKVLNQRQEELEADLKKREAIIVSKKEALLQE--KSHLENKK 851  
 QY 83 IADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISA----- 121  
 : | :: : | |:: |:: : : : : |  
 Db 852 LRSSQALNTDSLKISTR--NLLEQELSEKNVQLQTSTAEKTKISEQVEVLQEKDQLQ 909  
 QY 122 -----HPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFS 162  
 |::| | :| || : | : |:: : |  
 Db 910 KRRHDVDEKLKNGRVLSPREEHVLFLQLEEGIEALEAAIE---YRNESIQRQKSLRASPH 966  
 QY 163 NQFRDFFKYVE---LSTFDIASDAFATFKDLLT-----RHKVLVAD--- 200  
 | | :| || :| : | : : : ||| |  
 Db 967 NLSRGEANVLEKLACLSPVEIRTILFRYFNKVVNLRERKQQLYNEEMKMKVLERDNMV 1026  
 QY 201 -----FLEQNYDTI-----FEDYEKLLQS 219  
 | ||: : | :|| :|:  
 Db 1027 RELESALDHLKLQCDRRLTLQKEHEQKMQLLHHFKEQDGEIMETFKTYEDKIQQLEK 1086



Db 967 NLSRGEANVLEKLAQLSPVEIRTLFRYFNKVNNLREAERKQQLYNEEMKMKVLERDNMV 1026

Qy 201 -----FLEQNYDTI-----FEDYEKLLQS 219  
 | | : : | : | : | :

Db 1027 RELESALDHLKLQCDRRLTLQKKEHEQKMLLLHHFKEQDGEIMETFKTYEDKIQQLEK 1086

Qy 220 ENYVTKRQS-----LKLLGELILDRHNFAIM-----TKYISK 251  
 : | : | : : | | | : | :

Db 1087 DLYFYKKTSRDHKKKLKELVGEAI--RRQLAPSEYQEAGDGVLPKPEGGMLSEELKWASR 1144

Qy 252 PENLKLMMNLLRDKSPNIQFEAFHVFKVFPVSPHKTQPIVEILLKNQPKLIEFLSSF--- 308  
 ||: || : : : || : || : | : | ||

Db 1145 PESMKLSG---REREMDSS-----ASSLRTQPNPQKLWEDIPELPPHSSSLAPP 1190

Qy 309 -----QKERTDDEQFADEKKNYLIKQIR 330  
 | || | : | ||

Db 1191 SGHMLGNENKTETDDNQFTKSHSRLSSQIQ 1220

# RESULT 13

US-09-417-485D-6

; Sequence 6, Application US/09417485D

; Patent No. 6541202

; GENERAL INFORMATION:

; APPLICANT: Long, David M.

; APPLICANT: Metz, Anneke M.

; APPLICANT: Love, Ruschelle A.

; TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes

; FILE REFERENCE: 47714-5009-US

; CURRENT APPLICATION NUMBER: US/09/417,485D

; CURRENT FILING DATE: 2002-06-14

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 2184

; TYPE: PRT

; ORGANISM: Plasmodium falciparum

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (330)..(335)

; OTHER INFORMATION: Xaa at position 330 = Leu or Ile;

; OTHER INFORMATION: Xaa at position 335 = Asp or Gly.

US-09-417-485D-6

Query Match 6.6%; Score 113; DB 4; Length 2184;

Best Local Similarity 21.9%; Pred. No. 0.057;

Matches 77; Conservative 58; Mismatches 140; Indels 76; Gaps 17;

Qy 1 MKKMPLFSKSHKNPAEIV--KILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNE 58  
 : : : ||: | : | | : : : | : : || | | : |

Db 309 LPEIDFFSEDRKEKSSSVGYDXKKKNXSNIKRFHNKINRTKEEKKKKWN--KIIINRNNI 366

Qy 59 KEPPTAQAQLAQELYSSGLLVTLIAD-----LQLIDFEGKKDVTQIFNN----- 103  
 : | : : : | : | : : | | | : |

Db 367 LQHNT--TNKCKTFLFNKHIIFDKIENNNIPLFIYDLLNYIFKSDQTYFYHNNFIDEYKQ 424

Qy 104 ILRRQI--GTRSPTVEYI--SAHPHILFMLLK--GYEAPQIALRCGIMLRECIRHEPLA 156

Db 425 KICKQIKCSTKKN DISHIITSRKENHLFHVQKLENNYKHPNI-----NKQLRKTIL 476  
 QY 157 KIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTR-HKV-----L 197  
 Db 477 KYVY--NYFKEFINNVINTKFGKIYRKFFPRKHILNKIHKIFKIIRLQIIKKYRIINIRM 534  
 QY 198 VADFLEQN-YDTIFEDYE-----KLLQSENYVTKR-QSLKLLGELILDRHNFAIMT 246  
 Db 535 NRKFIKQKVYDTFFKNYDFLSFSFKTYKIINFMVYITKKCIPIKLLG----SKHNFKIFL 590  
 QY 247 KYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFASPHKTQPIVEILLKN 297  
 Db 591 KNVKK---FLLFNYKESFSLNQVMKNIKVKNIFQKKISKYNIKNRILLKN 637

RESULT 14

US-08-630-822A-70

; Sequence 70, Application US/08630822A

; Patent No. 5840695

; GENERAL INFORMATION:

; APPLICANT: FRANK, GLENN R.

; APPLICANT: HUNTER, SHIRLEY WU

; APPLICANT: WALLENFELS, LYNDIA

; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS

; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS

; NUMBER OF SEQUENCES: 107

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheridan Ross P.C.

; STREET: 1700 Lincoln Street, Suite 3500

; CITY: Denver

; STATE: Colorado

; COUNTRY: U.S.A.

; ZIP: 80203

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/630,822A

; FILING DATE: 11-APR-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: CONNELL, GARY J.

; REGISTRATION NUMBER: 32,020

; REFERENCE/DOCKET NUMBER: 2618-17-C3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 863-9700

; TELEFAX: (303) 863-0223

; INFORMATION FOR SEQ ID NO: 70:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 586 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FEATURE:

```

;      NAME/KEY:  Xaa = any amino acid
;      LOCATION:  379
US-08-630-822A-70

```

Query Match 6.2%; Score 105; DB 2; Length 586;  
Best Local Similarity 20.0%; Pred. No. 0.054;  
Matches 77; Conservative 54; Mismatches 136; Indels 118; Gaps 15;

Qy	22	KDNLAI <sup>LEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTTEAVAQLAQELYSSGLLVT</sup>	81
Db	205	KT <sup>KIEVIKEEERKIREERQEAREEEEEQRKQAEALNALNASSAAAEASS--AQEL-----</sup>	254
Qy	82	LIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALR	141
Db	255	LIDTAPVIDAEKTPKV-----ATSP-VESPLAPPEVLIM-----GAPK----	291
Qy	142	CGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATPKDLLTRHKVLVADF	201
Db	292	-----TPVATEVDKNADEVEFTK-KDLEVVEDALDTLSKDKNNLVIEKEVIKDI	339
Qy	202	LEQ-----NYDTIFEDYEKL--	216
Db	340	KEEIADYQEDVEELKEAIVAAEKP <sup>DEIKETKGAQRLLKXVNMKITKMDTVVQEIESKES</sup>	399
Qy	217	-----LQSENYVTKRQSL---KLLGELILDRHNFAI-MTKYISKPENLKLMMNLL--	262
Db	400	EKKAKTLPLEAPRSATETQELDVRKERGEILIDELMDAIKKVKVNPDENRLKLIENILGR	459
Qy	263	--RDKSPNIQFEAFHVFKVF-----VASPHKTQPIVEILLKNQPKLIEFLSSFQKER	312
Db	460	IDTDKDRHIKVE--DVLKVIDIVEKEDGIMSTKQ <sup>LDELVQLLKKEE--VIELEEKKEKQE</sup>	515
Qy	313	TDDEQFADEK <sup>NYLIKQIRDLKKTAP</sup>	337
Db	516	SQQKSFVPPSETLHLESSOOKSTVP	540

RESULT 15

US-09-005-069-70

; Sequence 70, Application US/09005069

; Patent No. 5932470

## ; GENERAL INFORMATION:

APPLICANT: FRANK, GLENN R.

; APPLICANT: HUNTER, SHIRLEY WU

; APPLICANT: WALLENFELS, LYNDIA

7 TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS

7 TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS

```
; NUMBER OF SEQUENCES: 107
```

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross P.C.

; STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

; STATE: Colorado

; COUNTRY: U.S.A.

ZIP: 80203

## ; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
```



```

;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/005,069
;      FILING DATE:
;      CLASSIFICATION:
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/630,822
;      FILING DATE:  11-APR-1996
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  CONNELL, GARY J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-17-C3
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;      INFORMATION FOR SEQ ID NO:  70:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  586 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
;      FEATURE:
;      NAME/KEY:  Xaa = any amino acid
;      LOCATION:  379
US-09-005-069-70

```

```

Query Match          6.2%;  Score 105;  DB 2;  Length 586;
Best Local Similarity 20.0%;  Pred. No. 0.054;
Matches  77;  Conservative  54;  Mismatches 136;  Indels 118;  Gaps  15;

```

```

Qy      22 KDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTTEAVAQLAQELYSSGLLVT 81
      | : ::::: | : : | : : | : | : || : |||
Db      205 KTKIEVIKEERKIREERQEAREEEEEQRKQAEALNASSAAAEASS--AQEL----- 254

Qy      82 LIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALR 141
      || : || | | | | | | | | : | | : | :
Db      255 LIDTAPVIDAEKTPKV-----ATSP-VESPLAPPEVLIM-----GAPK---- 291

Qy      142 CGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADF 201
      | : | : : : | | : | : | | : | : | : |
Db      292 -----TPVATEVDKNADEVFTK-KDLEVVEDALDTLSKDKNNLVIEKEVIKDI 339

Qy      202 LEQ-----NYDTIFEDYEKL-- 216
      | : | : : |
Db      340 KEEIADYQEDVEELKEAIVAAEKPKDEIKETKGAQRLLKXVNMITKMDTVVQEIESKES 399

Qy      217 -----LQSENYVTKRQSL---KLLGELILDRHNFAI-MTKYISKPENLKLMMNLL-- 262
      | : : | : | | | | : : | | : | : | : |
Db      400 EKKAKTLPLEAPRSATETQELDVRKERGEILIDELMDAIIKKVKNVPDENRLKLIENILGR 459

Qy      263 --RDKSPNIQFEAFHVFKVF-----VASPHKTQPIVEILLKNQPKLIEFLSSFQKER 312
      || : | : | | | : | : : | : | : | : |
Db      460 IDTDKDRHIKVE--DVLKVIDIVEKEDGIMSTKQLDELVQLLKKEE--VIELEEKKEKQE 515

Qy      313 TDDEQFADEKNYLIKQIRDLKKTAP 337

```

Db           : : |       | :       | | |  
516 SQQKSFVPPSETLHLESSQQKSTVP 540

Search completed: January 7, 2004, 16:45:03  
Job time : 29 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 16:44:17 ; Search time 44 Seconds  
(without alignments)  
1215.701 Million cell updates/sec

Title: US-10-088-872-2  
Perfect score: 1704  
Sequence: 1 MKKMPLFSKSHKNPAEIVKI.....FADEKNYLIKQIRDLKKTAP 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result	% Query					Description
	No.	Score	Match Length DB	ID		
1	1704	100.0	337	21	AA94247	Human calcium bind
2	1704	100.0	337	22	AAM39078	Human polypeptide
3	1704	100.0	337	22	AAB82090	Human Acute Neuron
4	1466	86.0	289	22	AAB94139	Human protein sequ
5	1381	81.0	341	22	AAB48970	Human ANIC-BP (acu
6	1381	81.0	496	22	AAE10858	Gal4-human ANIC-BP
7	1381	81.0	552	22	AAE10859	LexA-human ANIC-BP
8	1376	80.8	341	21	AA94248	Mouse calcium bind
9	1354	79.5	354	22	ABG23844	Novel human diagno
10	1297.5	76.1	350	22	AAB20387	Human acute neuron
11	1162	68.2	237	22	AAM40864	Human polypeptide
12	1111	65.2	339	22	ABB60392	Drosophila melanog
13	1109	65.1	339	21	AA94249	Drosophila calcium
14	1063.5	62.4	377	21	AA94250	C. elegans yeast-1
15	716.5	42.0	343	21	AAG45273	Arabidopsis thalia
16	689.5	40.5	300	21	AAG23886	Arabidopsis thalia
17	685.5	40.2	400	21	AAG51052	Arabidopsis thalia
18	685.5	40.2	504	21	AAG51051	Arabidopsis thalia
19	685	40.2	300	21	AAG30714	Arabidopsis thalia
20	685	40.2	300	21	AAG45274	Arabidopsis thalia
21	685	40.2	305	21	AAG30713	Arabidopsis thalia
22	684.5	40.2	326	21	AAG51053	Arabidopsis thalia
23	675.5	39.6	290	21	AAG23887	Arabidopsis thalia
24	671.5	39.4	345	21	AAG05089	Arabidopsis thalia
25	638.5	37.5	320	21	AAG05090	Arabidopsis thalia
26	539.5	31.7	213	21	AAG23888	Arabidopsis thalia
27	533	31.3	213	21	AAG30715	Arabidopsis thalia
28	533	31.3	213	21	AAG45275	Arabidopsis thalia
29	478.5	28.1	197	21	AAG05091	Arabidopsis thalia
30	467.5	27.4	154	21	AAG41151	Zea mays protein f
31	453.5	26.6	148	21	AAG41152	Zea mays protein f
32	438.5	25.7	139	21	AAG41153	Zea mays protein f
33	250.5	14.7	236	23	ABP02921	Human ORFX protein
34	241	14.1	639	22	ABG25372	Novel human diagno
35	227.5	13.4	135	23	ABP34081	Human ORF3054 prot
36	226.5	13.3	383	22	ABG23843	Novel human diagno
37	125	7.3	660	22	ABB30817	Peptide #3468 enco
38	125	7.3	660	23	ABG38772	Human peptide enco
39	117.5	6.9	709	23	ABG70293	Human novel polype
40	114.5	6.7	833	21	AAB42353	Human ORFX ORF2117
41	113.5	6.7	1279	23	ABG70787	Human kinesin-rela
42	113.5	6.7	1279	23	ABB80078	Human kinesin moto
43	113.5	6.7	1279	24	ABG72397	Human partial kine
44	113	6.6	2184	22	AAE00425	P. falciparum telo
45	111.5	6.5	725	18	AAW39165	Human RHAMM protei

# ALIGNMENTS

RESULT 1

AAY94247

ID AAY94247 standard; protein; 337 AA.

XX

AC AAY94247;

XX

DT 10-AUG-2000 (first entry)

XX

DE Human calcium binding protein hCBP.

XX

KW Human; calcium binding protein; cancer; inflammation; CBP;  
KW reproductive disorder; autoimmune disorder; developmental disorder;  
KW seizure disorder; immune disorder; infection.

XX

OS Homo sapiens.

XX

PN WO200029580-A1.

XX

PD 25-MAY-2000.

XX

PF 12-NOV-1999; 99WO-US27027.

XX

PR 13-NOV-1998; 98US-0190965.

XX

PA (INCY-) INCYTE PHARM INC.

XX

PI Tang YT, Guegler KJ, Corley NC, Gorgone GA;

XX

DR WPI; 2000-387793/33.

DR N-PSDB; AAA27332.

XX

PT Human hCBP protein, and the nucleic acid encoding it, useful for e.g.  
PT diagnosis, prevention and treatment of cancers, immune, developmental  
PT or reproductive disorders -

XX

PS Claim 1; Fig 1; 72pp; English.

XX

CC The present sequence is the human calcium binding protein hCBP. It  
CC was obtained by screening a coronary artery smooth muscle cDNA library,  
CC from which five overlapping nucleic acids were isolated, sequenced and  
CC expressed to give the protein. The protein and the gene encoding it are  
CC useful for the diagnosis and treatment of the following types of  
CC disorder: cancers (such as adenocarcinomas), reproductive disorders  
CC (such as infertility, ovulatory defects, endometriosis, disruptions of  
CC the oestrus and menstrual cycles, polycystic ovary syndrome and ovarian  
CC hyperstimulation), autoimmune disorders (such as benign prostatic  
CC hyperplasia and prostatitis), developmental disorders (such as  
CC Cushing's syndrome, muscular dystrophy and gonadal dysgenesis),  
CC hereditary neuropathies, seizure disorders, immune disorders (such as  
CC AIDS, allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's  
CC disease, diabetes, Graves' disease, multiple sclerosis, psoriasis,  
CC rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative  
CC colitis), and viral, bacterial, fungal, parasitic, protozoal and  
CC helminthic infections.

XX

SQ Sequence 337 AA;

Query Match 100.0%; Score 1704; DB 21; Length 337;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-146;  
 Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
        |||
Db      1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60

Qy     61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
        |||
Db     61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120

Qy    121 AHPHILFMLLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
        |||
Db    121 AHPHILFMLLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180

Qy    181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
        |||
Db    181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240

Qy    241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQPK 300
        |||
Db    241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQPK 300

Qy    301 LIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAP 337
        |||
Db    301 LIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAP 337
  
```

## RESULT 2

AAM39078

ID AAM39078 standard; Protein; 337 AA.

XX

AC AAM39078;

XX

DT 22-OCT-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 2223.

XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX

OS Homo sapiens.

XX

PN WO200153312-A1.

XX

PD 26-JUL-2001.

XX

PF 26-DEC-2000; 2000WO-US34263.

XX

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.



Db 181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240

QY 241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQPK 300  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQPK 300  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 301 LIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAP 337  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 301 LIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAP 337

RESULT 3

AAB82090

ID AAB82090 standard; Protein; 337 AA.

XX

AC AAB82090;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human Acute Neuronal Induced Calcium Binding Protein, ANIC-BP.

XX

KW Human; cerebroprotective; neuroprotective; vulnerary; vaccine;  
 KW gene therapy; Acute Neuronal Induced Calcium Binding Protein; ANIC-BP;  
 KW stroke; acute head trauma; multiple sclerosis; spinal cord injury.

XX

OS Homo sapiens.

XX

PN WO200123552-A1.

XX

PD 05-APR-2001.

XX

PF 18-SEP-2000; 2000WO-EP09132.

XX

PR 24-SEP-1999; 99EP-0118848.

XX

PA (MERE ) MERCK PATENT GMBH.

XX

PI Den Daas I, Duecker K;

XX

DR WPI; 2001-308142/32.

DR

N-PSDB; AAF86462.

XX

PT Novel human acute neuronal induced calcium binding polypeptide, and  
 PT polynucleotides encoding them useful for diagnosing or treating stroke,  
 PT acute head trauma, multiple sclerosis and spinal cord injury -

XX

PS Claim 1; Page 41-42; 45pp; English.

XX

CC The present sequence is the protein sequence for human Acute Neuronal  
 CC Induced Calcium Binding Protein (ANIC-BP). ANIC-BP coding sequence and  
 CC protein are useful for treating stroke, acute head trauma, multiple  
 CC sclerosis and spinal cord injury. ANIC-BP coding sequence and protein  
 CC are also useful as vaccines for inducing an immunological response in a  
 CC mammal.

XX

SQ Sequence 337 AA;



Query Match 100.0%; Score 1704; DB 22; Length 337;  
Best Local Similarity 100.0%; Pred. No. 1.3e-146;  
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60

Qy     61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120

Qy    121 AHPHILFMLLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AHPHILFMLLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180

Qy    181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240

Qy    241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVVASPHKTQPIVEILLKNQPK 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVVASPHKTQPIVEILLKNQPK 300

Qy    301 LIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAP 337
        ||||||||||||||||||||||||||||||||||||
Db    301 LIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAP 337
```

RESULT 4

AAB94139

ID AAB94139 standard; Protein; 289 AA.

XX

AC AAB94139;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:14408.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-0116126.

XX

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX

PA (HELI-) HELIX RES INST.

XX



QY 289 PIVEILLKNQPKLIEFLSSQKERTDDEQFADEKNYLIKQIRDLKKTAP 337  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 PIVEILLKNQPKLIEFLSSQKERTDDEQFADEKNYLIKQIRDLKKTAP 289

RESULT 5

AAB48970

ID AAB48970 standard; Protein; 341 AA.

XX

AC AAB48970;

XX

DT 27-MAR-2001 (first entry)

XX

DE Human ANIC-BP (acute neuronal induced calcium-binding protein).

XX

KW Human; acute neuronal induced calcium-binding protein; ANIC-BP;

KW Mo25 homologue; HymA homologue; drug screening; stroke;

KW acute head trauma; multiple sclerosis; spinal cord injury; vaccine;

KW cerebroprotective; neuroprotective.

XX

OS Homo sapiens.

XX

PN WO200078947-A1.

XX

PD 28-DEC-2000.

XX

PF 14-JUN-2000; 2000WO-EP05457.

XX

PR 22-JUN-1999; 99EP-0112024.

XX

PA (MERE ) MERCK PATENT GMBH.

XX

PI Den Daas I, Fischer V, Seyfried C, Von Melchner L;

XX

DR WPI; 2001-102721/11.

DR N-PSDB; AAC91772.

XX

PT Novel acute neuronal induced calcium binding protein, useful for

PT treating acute head trauma, stroke, multiple sclerosis and spinal cord

PT injury -

XX

PS Claim 2; Page 37; 50pp; English.

XX

CC The invention relates to human acute neuronal induced calcium-binding

CC protein (ANIC-BP) and to nucleic acid encoding it. The invention

CC also relates to expression systems and recombinant host cells comprising

CC ANIC-BP DNA, the recombinant production of ANIC-BP, antibodies specific

CC for ANIC-BP, fusion proteins comprising ANIC-BP and an immunoglobulin

CC Fc region, and methods of screening for modulators of ANIC-BP function.

CC ANIC-BP has homology and structural similarity to HymA and Mo25 proteins.

CC ANIC-BP proteins and nucleotides are useful for treating stroke and

CC acute head trauma, multiple sclerosis and spinal cord injury. ANIC-BP

CC proteins are useful in screening assays, for identifying membrane bound

CC or soluble receptors, and also in vaccines. ANIC-BP nucleotides are

CC useful as diagnostic reagents, as tools for tissue expression studies,

CC for chromosome localisation studies, as genetic vaccines, and in

CC the generation of transgenic animals. The present sequence represents

CC human ANIC-BP.  
XX  
SQ Sequence 341 AA;

Query Match 81.0%; Score 1381; DB 22; Length 341;  
Best Local Similarity 81.0%; Pred. No. 3.2e-117;  
Matches 273; Conservative 31; Mismatches 29; Indels 4; Gaps 2;

```
Qy      4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
      || | |||:|:| ||::|:| || | :|:| |||:| ||||| |||||
Db      1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy     60 EPPTTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
      || ||||| |||:|:| ||| ||||| ||||| ||||| |||||
Db     61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy    120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
      : ||||| |||:| ||| ||||| ||||| ||||| ||||| |||||
Db    121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy    180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
      ||||| ||||| |||:| |||:| || | :||| ||||| ||||| ||||| |||||
Db    181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLDR 240

Qy    240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQP 299
      || | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVVFVANPNKTQPIILDILLKNQA 300

Qy    300 KLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAA 336
      ||||| || | :|:| ||| || | :||| |||:|
Db    301 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRDLKRPA 337
```

RESULT 6

AAE10858

ID AAE10858 standard; Protein; 496 AA.

XX

AC AAE10858;

XX

DT 18-DEC-2001 (first entry)

XX

DE Gal4-human ANIC-BP-1 fusion protein.

XX

KW Human; acute neuronal induced calcium binding protein type 1 ligand;

KW ANIC-BP-1; human disease; stroke; head trauma; multiple sclerosis;

KW Parkinson's disease; Alzheimer's disease; spinal cord injury; vaccine;

KW gene therapy; fusion protein; Gal4 protein.

XX

OS Chimeric - Homo sapiens.

OS Chimeric - Unidentified.

XX

PN WO200170771-A2.

XX

PD 27-SEP-2001.

XX

PF 20-MAR-2001; 2001WO-EP03149.

XX

PR 21-MAR-2000; 2000EP-0106110.  
XX  
PA (MERE ) MERCK PATENT GMBH.  
XX  
PI Den Daas I, Duecker K, Hock B;  
XX  
DR WPI; 2001-607519/69.  
XX  
PT Novel acute neuronal induced calcium binding protein type 1 ligand  
PT polypeptides, useful in the treatment of stroke, head trauma, multiple  
PT sclerosis, Parkinson's disease, Alzheimer's disease and spinal cord  
PT injury -  
XX

PS Disclosure; Page 42-44; 46pp; English.  
XX

CC The invention relates to human acute neuronal induced calcium binding  
CC protein type 1 (ANIC-BP-1) ligand polypeptides and polynucleotides.  
CC Sequences of the invention are useful for treating human diseases  
CC including stroke, head trauma, multiple sclerosis, Parkinson's disease,  
CC Alzheimer's disease and spinal cord injury. They are also useful as  
CC vaccines. ANIC-BP-1 ligands are useful for identifying membrane bound  
CC soluble receptors. Polynucleotides of the invention are useful as  
CC diagnostic reagents, for chromosome localization studies, and as  
CC valuable tools for tissue expression studies. They are also useful in  
CC gene therapy. The present sequence is Gal4-human ANIC-BP-1 fusion  
CC protein comprising the Gal4 protein and a C-terminally linked human  
CC ANIC-BP-1 protein.  
XX

SQ Sequence 496 AA;

Query Match 81.0%; Score 1381; DB 22; Length 496;  
Best Local Similarity 81.0%; Pred. No. 5.2e-117;  
Matches 273; Conservative 31; Mismatches 29; Indels 4; Gaps 2;

Qy 4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59  
|| | |||:|:| |||:|:| ||| | | :|:| |||:| ||| ||| |||  
Db 156 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 215  
  
Qy 60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTEVEYI 119  
|| ||||| |||:| ||| |||:| ||||| ||||| ||||| ||||| |||||  
Db 216 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTEVEYI 275  
  
Qy 120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179  
: ||||| |||:| ||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 276 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 335  
  
Qy 180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239  
||| ||||| ||||| |||:| ||| |||:| ||||| ||||| ||||| |||||  
Db 336 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 395  
  
Qy 240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFASPHKTQPIVEILLKNQP 299  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 396 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVVFVANPNKTQPILDILLKNQA 455  
  
Qy 300 KLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTA 336  
||| ||| |||:| ||| ||| |||:| ||| |||:| ||| |||:|  
Db 456 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRDLKRPA 492

RESULT 7

AAE10859

ID AAE10859 standard; Protein; 552 AA.

XX

AC AAE10859;

XX

DT 18-DEC-2001 (first entry)

XX

DE LexA-human ANIC-BP-1 fusion protein.

XX

KW Human; acute neuronal induced calcium binding protein type 1 ligand;

KW ANIC-BP-1; human disease; stroke; head trauma; multiple sclerosis;

KW Parkinson's disease; Alzheimer's disease; spinal cord injury; vaccine;

KW gene therapy; fusion protein; LexA protein.

XX

OS Chimeric - Homo sapiens.

OS Chimeric - Unidentified.

XX

FH Key Location/Qualifiers

FT Region 1..202

FT /note= "LexA protein"

FT Region 203..552

FT /note= "Human ANIC-BP-1 protein"

XX

PN WO200170771-A2.

XX

PD 27-SEP-2001.

XX

PF 20-MAR-2001; 2001WO-EP03149.

XX

PR 21-MAR-2000; 2000EP-0106110.

XX

PA (MERE ) MERCK PATENT GMBH.

XX

PI Den Daas I, Duecker K, Hock B;

XX

DR WPI; 2001-607519/69.

XX

PT Novel acute neuronal induced calcium binding protein type 1 ligand

PT polypeptides, useful in the treatment of stroke, head trauma, multiple

PT sclerosis, Parkinson's disease, Alzheimer's disease and spinal cord

PT injury -

XX

PS Disclosure; Page 44-46; 46pp; English.

XX

CC The invention relates to human acute neuronal induced calcium binding  
CC protein type 1 (ANIC-BP-1) ligand polypeptides and polynucleotides.

CC Sequences of the invention are useful for treating human diseases

CC including stroke, head trauma, multiple sclerosis, Parkinson's disease,

CC Alzheimer's disease and spinal cord injury. They are also useful as

CC vaccines. ANIC-BP-1 ligands are useful for identifying membrane bound

CC soluble receptors. Polynucleotides of the invention are useful as

CC diagnostic reagents, for chromosome localization studies, and as

CC valuable tools for tissue expression studies. They are also useful in

CC gene therapy. The present sequence is LexA-human ANIC-BP-1 fusion

CC protein comprising the LexA protein and a C-terminally linked human  
CC ANIC-BP-1 protein.  
XX  
SQ Sequence 552 AA;

Query Match 81.0%; Score 1381; DB 22; Length 552;  
Best Local Similarity 81.0%; Pred. No. 6e-117;  
Matches 273; Conservative 31; Mismatches 29; Indels 4; Gaps 2;

```
Qy      4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
      || | |||:|:| |||:::|:| ||| :|:| |||:| ||||| |||||
Db     212 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 271

Qy     60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
      || ||||| ||||| |||:|:| |||:|:| ||||| ||||| |||||
Db     272 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 331

Qy    120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
      : ||||| |||:|:| ||| ||||| ||||| ||||| ||||| |||||
Db    332 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 391

Qy    180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
      ||||| ||||| |||:|:| ||| | :||| ||||| ||||| ||||| |||||
Db    392 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLDDR 451

Qy    240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQP 299
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    452 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVVFVANPNKTQPILDILLKNQA 511

Qy    300 KLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLLKKA 336
      ||||| || :||:| ||| ||| ||:| ||||| :|
Db    512 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRDLLKRP 548
```

RESULT 8

AA94248

ID AAY94248 standard; protein; 341 AA.

XX

AC AAY94248;

XX

DT 10-AUG-2000 (first entry)

XX

DE Mouse calcium binding protein MO25.

XX

KW Mouse; calcium binding protein; cancer; inflammation; MO25; CBP;  
KW reproductive disorder; autoimmune disorder; developmental disorder;  
KW seizure disorder; immune disorder; infection.

XX

OS Mus sp.

XX

PN WO200029580-A1.

XX

PD 25-MAY-2000.

XX

PF 12-NOV-1999; 99WO-US27027.

XX

PR 13-NOV-1998; 98US-0190965.

XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Tang YT, Guegler KJ, Corley NC, Gorgone GA;  
XX  
DR WPI; 2000-387793/33.  
XX  
PT Human hCBP protein, and the nucleic acid encoding it, useful for e.g.  
PT diagnosis, prevention and treatment of cancers, immune, developmental  
PT or reproductive disorders -  
XX  
PS Disclosure; Page 66-67; 72pp; English.  
XX  
CC The present sequence is the mouse calcium binding protein MO25. It  
CC was used in a sequence alignment to identify human calcium binding  
CC protein hCBP. The hCBP protein and the gene encoding it are  
CC useful for the diagnosis and treatment of the following types of  
CC disorder: cancers (such as adenocarcinomas), reproductive disorders  
CC (such as infertility, ovulatory defects, endometriosis, disruptions of  
CC the oestrus and menstrual cycles, polycystic ovary syndrome and ovarian  
CC hyperstimulation), autoimmune disorders (such as benign prostatic  
CC hyperplasia and prostatitis), developmental disorders (such as  
CC Cushing's syndrome, muscular dystrophy and gonadal dysgenesis),  
CC hereditary neuropathies, seizure disorders, immune disorders (such as  
CC AIDS, allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's  
CC disease, diabetes, Graves' disease, multiple sclerosis, psoriasis,  
CC rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative  
CC colitis), and viral, bacterial, fungal, parasitic, protozoal and  
CC helminthic infections.  
XX  
SQ Sequence 341 AA;

Query Match 80.8%; Score 1376; DB 21; Length 341;  
Best Local Similarity 80.7%; Pred. No. 9e-117;  
Matches 272; Conservative 32; Mismatches 29; Indels 4; Gaps 2;

Qy	4	MPL-FSKSHKNPAEIVKILKDNLAILEKQ--DKKTDKASEEVSKSLQAMKEILCGTNEK	59
		:     :           :	
Db	1	MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK	60
Qy	60	EPPTTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTEVEYI	119
		:           :	
Db	61	EPQTEAVAQLAQELYNSGLLGLTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTEVEYI	120
Qy	120	SAHPHILFMLLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI	179
		:                 :   :	
Db	121	CTQQNILFMLLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI	180
Qy	180	ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR	239
		:   :       :	
Db	181	ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLDDR	240
Qy	240	HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQP	299
		:   :       :	
Db	241	HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVVFVANPNKTQPIILDILLKNQT	300
Qy	300	KLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAA	336



Db

||||| || :||:|||| ||| ||:||||:|:| |  
301 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRNLKRAA 337

RESULT 9

ABG23844

ID ABG23844 standard; Protein; 354 AA.

XX

AC ABG23844;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #23835.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS88031.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX

PS Claim 20; SEQ ID No 54203; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity



XX  
PD 12-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-EP09475.  
XX  
PR 04-OCT-1999; 99EP-0119113.  
XX  
PA (MERE ) MERCK PATENT GMBH.  
XX  
PI Duecker K, Den Daas I;  
XX  
DR WPI; 2001-266306/27.  
DR N-PSDB; AAF30688.  
XX  
PT Novel human acute neuronal induced calcium-binding protein like protein  
PT splice variant, useful for treating stroke, acute head trauma,  
PT Parkinson's disease, Alzheimer's disease multiple sclerosis, spinal  
PT cord injury -  
XX  
PS Claim 2; Page 44-45; 49pp; English.  
XX  
CC The present sequence is that of a novel human acute neuronal induced  
CC calcium binding protein-like protein splice variant, ANIC-NP-1B.  
CC The protein shows homology to other members of the calcium binding  
CC protein family, including ANIC-BP, a protein discovered by mRNA  
CC differential display that is upregulated in a rat model of head  
CC trauma. ANIC-BP and ANIC-BP-1B differ in their C-terminal portions.  
CC The variant protein could serve as a novel drug target. The  
CC invention provides ANIC-BP-1B polynucleotides (see AAF30688) and  
CC polypeptides, expression vectors, host cells and antibodies, as  
CC well as methods for producing the protein and for treating or  
CC preventing disorders associated with expression of the protein by  
CC inhibiting or activating the action of ANIC-BP-1B. Diseases that  
CC may be treated include stroke and acute head trauma, Parkinson's  
CC disease, Alzheimer's disease, multiple sclerosis and spinal cord  
CC injury. The polynucleotides and polypeptides can also be used in  
CC diagnostic assays and in vaccines, and to identify agonists and  
CC antagonists useful for treating conditions associated with  
CC ANIC-BP-1B imbalance.  
XX  
SQ Sequence 350 AA;

Query Match 76.1%; Score 1297.5; DB 22; Length 350;  
Best Local Similarity 76.0%; Pred. No. 1.3e-109;  
Matches 263; Conservative 32; Mismatches 38; Indels 13; Gaps 3;

Qy 4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59  
|| | ||||:|:| ||:::|:| | | :|:| ||||:| ||||  
Db 1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60  
Qy 60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTEVEYI 119  
|| ||||| ||||:| ||:| ||||| ||||| ||||| ||||:| ||||  
Db 61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTEVEYI 120  
Qy 120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179  
: ||||| ||:| || ||||| ||||| ||||:| || ||:|:| ||||  
Db 121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFRYVEMSTFDI 180







Db 181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLENSENYVTRRQSLKLLGELLDDR 240  
 Qy 240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQP 299  
 ||| :||:||||:|||||||:|:| | ||||| |||||:|:| :||:||||:|  
 Db 241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVVFVANPNKPKPILDILLRNQT 300  
 Qy 300 KLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLK 333  
 ||:|:||:| :|:| ||| ||| |||||:|:|  
 Db 301 KLVDFLTNFHTDRSEDEQFNDEKAYLIKQIKELK 334

RESULT 13

AA94249

ID AAY94249 standard; protein; 339 AA.

XX

AC AAY94249;

XX

DT 10-AUG-2000 (first entry)

XX

DE Drosophila calcium binding protein DMO25.

XX

KW Drosophila; calcium binding protein; cancer; inflammation; DMO25; CBP;  
 KW reproductive disorder; autoimmune disorder; developmental disorder;  
 KW seizure disorder; immune disorder; infection.

XX

OS Drosophila melanogaster.

XX

PN WO200029580-A1.

XX

PD 25-MAY-2000.

XX

PF 12-NOV-1999; 99WO-US27027.

XX

PR 13-NOV-1998; 98US-0190965.

XX

PA (INCY-) INCYTE PHARM INC.

XX

PI Tang YT, Guegler KJ, Corley NC, Gorgone GA;

XX

DR WPI; 2000-387793/33.

XX

PT Human hCBP protein, and the nucleic acid encoding it, useful for e.g.  
 PT diagnosis, prevention and treatment of cancers, immune, developmental  
 PT or reproductive disorders -

XX

PS Disclosure; Page 67-68; 72pp; English.

XX

CC The present sequence is the Drosophila calcium binding protein DMO25. It  
 CC was used in a sequence alignment to identify human calcium binding  
 CC protein hCBP. The hCBP protein and the gene encoding it are  
 CC useful for the diagnosis and treatment of the following types of  
 CC disorder: cancers (such as adenocarcinomas), reproductive disorders  
 CC (such as infertility, ovulatory defects, endometriosis, disruptions of  
 CC the oestrus and menstrual cycles, polycystic ovary syndrome and ovarian  
 CC hyperstimulation), autoimmune disorders (such as benign prostatic  
 CC hyperplasia and prostatitis), developmental disorders (such as  
 CC Cushing's syndrome, muscular dystrophy and gonadal dysgenesis),

CC hereditary neuropathies, seizure disorders, immune disorders (such as  
 CC AIDS, allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's  
 CC disease, diabetes, Graves' disease, multiple sclerosis, psoriasis,  
 CC rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative  
 CC colitis), and viral, bacterial, fungal, parasitic, protozoal and  
 CC helminthic infections.

XX

SQ Sequence 339 AA;

Query Match 65.1%; Score 1109; DB 21; Length 339;  
 Best Local Similarity 65.0%; Pred. No. 1.8e-92;  
 Matches 217; Conservative 59; Mismatches 54; Indels 4; Gaps 3;

Qy 4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTKDASEEVSKSLQAMKEILCGTNEKEPPT 63  
 |||| | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 1 MPLFGKSQKSPVELVKSLEAINALEAGDRKVEKAQEDVSKNLVSIKNMLHGSSDAEPPA 60  
 Qy 64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122  
 : ||||:||||: | | : | | : | | |||| | | ||||:|||||||  
 Db 61 DYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTVEYICTK 120  
 Qy 123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180  
 | || | | : | | | : | | | | | | | : | | | | | : | : | | | : | | | |  
 Db 121 PEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180  
 Qy 181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLQSENYVTKRQSLKLLGELILDR 239  
 ||||:||||:||||||: | | : | | | | : | : | | | | | | | | | | | | | |  
 Db 181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLNSENENYVTRRQSLKLLGELLDR 240  
 Qy 240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299  
 || | : | | : | | | | | | | | : | | | | | | | | | | | | | | | | | | |  
 Db 241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILLRNQT 300  
 Qy 300 KLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLK 333  
 | : : | | : | : | : | | | | | | | | : | |  
 Db 301 KLVDFLT NFHTDRSEDEQFNDEKAYLIKQIKELK 334

# RESULT 14

AA94250

ID AAY94250 standard; protein; 377 AA.

XX

AC AAY94250;

XX

DT 10-AUG-2000 (first entry)

XX

DE C. elegans yeast-like calcium binding protein.

XX

KW Calcium binding protein; cancer; inflammation; yeast-like CBP; CBP;  
 KW reproductive disorder; autoimmune disorder; developmental disorder;  
 KW seizure disorder; immune disorder; infection.

XX

OS Caenorhabditis elegans.

XX

PN WO200029580-A1.

XX

PD 25-MAY-2000.



XX  
PF 12-NOV-1999; 99WO-US27027.  
XX  
PR 13-NOV-1998; 98US-0190965.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Tang YT, Guegler KJ, Corley NC, Gorgone GA;  
XX  
DR WPI; 2000-387793/33.  
XX  
PT Human hCBP protein, and the nucleic acid encoding it, useful for e.g.  
PT diagnosis, prevention and treatment of cancers, immune, developmental  
PT or reproductive disorders -  
XX  
PS Disclosure; Page 68-69; 72pp; English.  
XX  
CC The present sequence is the *C. elegans* yeast-like CBP. It  
CC was used in a sequence alignment to identify human calcium binding  
CC protein hCBP. The hCBP protein and the gene encoding it are  
CC useful for the diagnosis and treatment of the following types of  
CC disorder: cancers (such as adenocarcinomas), reproductive disorders  
CC (such as infertility, ovulatory defects, endometriosis, disruptions of  
CC the oestrus and menstrual cycles, polycystic ovary syndrome and ovarian  
CC hyperstimulation), autoimmune disorders (such as benign prostatic  
CC hyperplasia and prostatitis), developmental disorders (such as  
CC Cushing's syndrome, muscular dystrophy and gonadal dysgenesis),  
CC hereditary neuropathies, seizure disorders, immune disorders (such as  
CC AIDS, allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's  
CC disease, diabetes, Graves' disease, multiple sclerosis, psoriasis,  
CC rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative  
CC colitis), and viral, bacterial, fungal, parasitic, protozoal and  
CC helminthic infections.  
XX  
SQ Sequence 377 AA;

Query Match 62.4%; Score 1063.5; DB 21; Length 377;  
Best Local Similarity 60.5%; Pred. No. 2.8e-88;  
Matches 211; Conservative 53; Mismatches 68; Indels 17; Gaps 3;

QY 4 MP-LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAM 49  
|| || ||||:|::|| |:: | ||:| || || :|||:: :  
Db 1 MPLLFQKSHKSPADVVKTLREVLTLTDKLPKPKLDKGNIQSDKKYDKALDEVSKNVAMI 60

QY 50 KEILCGTNEKEPPTE---AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR 106  
| : | : || :| |||||:|:: :| || | :|| |||| ||||:|  
Db 61 KSFIYGNDSEPSSEHVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVQGFNNLLR 120

QY 107 RQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFR 166  
|||||||:| | || |::|| | || | ||:| ||||:|:|  
Db 121 RQIGTRSPTVEYILGARPEILIQLVQGYSPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180

QY 167 DFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKR 226  
|| ||: |||:||||:|::| |||| ::|:|: |||| | : || |:|:|:|  
Db 181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIEFLDSNYDTFFAQYQNLNLSKNYVTRR 240

QY 227 QSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVAPSK 286

Db 241 QSLKLLGELLDRHNFNTMTKYISNPDNLRMMELLRDKSRNIQYEAFHVFKVVFVANPNK 300  
 QY 287 TQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKT 335  
 Db 301 PKPISDILNRNREKLVEFLSEFHNDRTDDEQFNDEKAYLIKQIQEMKSS 349

RESULT 15

AAG45273

ID AAG45273 standard; Protein; 343 AA.

XX

AC AAG45273;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 56816.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.

PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.

PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.

PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 42.0%; Score 716.5; DB 21; Length 343;  
 Best Local Similarity 42.9%; Pred. No. 9.3e-57;  
 Matches 144; Conservative 78; Mismatches 105; Indels 9; Gaps 3;

QY 6 LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAMKEILCGTNE 58  
 || : ||:||: :| | :: :| :| :|||:: :| || :|  
 Db 4 LFKSKPRTPADIVRQTRDLLLYADRSNSFPDLRESKREEKMVELSKSIRDCLKLILYGNSE 63  
 QY 59 KEPPTAEVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEY 118  
 || || ||| || : : | | :| :| ||| :| :| :| :|  
 Db 64 AEPVAEACAQLTQEFFKADTLRRLLTSLPNLNLEARKDATQVVANLQRQQVNSRLIAADY 123  
 QY 119 ISAPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFD 178  
 : :: :| :| :|| | ||||| :| :| :| :| :| :| :|  
 Db 124 LESNIDLMDFLVDGFENTDMALHYGTMFRECIRHQIVAKYVLDSEHVKKFFYYIQLPNFD 183  
 QY 179 IASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDY-EKLLQSENYVTKRQSLKLLGELIL 237

Db	184	IAADAAATFKELLTRHKSTVAEFLIKNEDWFFADYNSKLLLESTNYITRRQAIKLLGDILL	243
Qy	238	DRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKN	297
Db	244	DRSNSAVMTKYVSSMDNLRILMNLLRESSKTIQIEAFHVFKLFVANQNKPSDIANILVAN	303
Qy	298	QPKLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLK	333
Db	304	RNKLLRLLADIKPDK-EDERFDADKAQVVREIANLK	338

Search completed: January 7, 2004, 16:47:07  
Job time : 55 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 16:46:15 ; Search time 35 Seconds  
(without alignments)  
1940.251 Million cell updates/sec

Title: US-10-088-872-2  
Perfect score: 1704  
Sequence: 1 MKKMPLFSKSHKNPAEIVKI.....FADEKNYLIKQIRDLKKTAP 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query					
No.	Score	Match	Length	DB	ID	Description	

1	1704	100.0	337	15	US-10-025-730-1	Sequence 1, Appli
2	1381	81.0	496	12	US-10-239-079-5	Sequence 5, Appli
3	1381	81.0	552	12	US-10-239-079-6	Sequence 6, Appli
4	1376	80.8	341	15	US-10-025-730-3	Sequence 3, Appli
5	1109	65.1	339	15	US-10-025-730-4	Sequence 4, Appli
6	1063.5	62.4	377	15	US-10-025-730-5	Sequence 5, Appli
7	128.5	7.5	820	12	US-10-029-386-32324	Sequence 32324, A
8	128.5	7.5	3878	12	US-10-080-608A-11	Sequence 11, Appl
9	128.5	7.5	3899	15	US-10-171-311-4	Sequence 4, Appli
10	128.5	7.5	3907	15	US-10-171-311-2	Sequence 2, Appli
11	128.5	7.5	3911	12	US-10-370-685-100	Sequence 100, App
12	128.5	7.5	3917	15	US-10-171-311-8	Sequence 8, Appli
13	128.5	7.5	3925	15	US-10-171-311-6	Sequence 6, Appli
14	125	7.3	660	9	US-09-864-761-47959	Sequence 47959, A
15	116.5	6.8	709	12	US-10-023-634-18	Sequence 18, Appl
16	116.5	6.8	709	12	US-10-023-634-77	Sequence 77, Appl
17	116.5	6.8	724	12	US-10-023-634-80	Sequence 80, Appl
18	113	6.6	2184	12	US-10-304-095-6	Sequence 6, Appli
19	111.5	6.5	725	12	US-10-023-634-79	Sequence 79, Appl
20	108	6.3	769	12	US-10-032-585-7117	Sequence 7117, Ap
21	106.5	6.2	709	12	US-10-256-250-15	Sequence 15, Appl
22	106.5	6.2	709	12	US-10-023-634-14	Sequence 14, Appl
23	106.5	6.2	725	11	US-09-978-309A-47	Sequence 47, Appl
24	106.5	6.2	725	12	US-10-256-250-14	Sequence 14, Appl
25	106.5	6.2	725	12	US-10-023-634-78	Sequence 78, Appl
26	106	6.2	646	12	US-10-023-634-16	Sequence 16, Appl
27	106	6.2	1295	12	US-10-369-493-6440	Sequence 6440, Ap
28	105.5	6.2	1091	12	US-10-369-493-6328	Sequence 6328, Ap
29	105	6.2	586	14	US-10-071-751-30	Sequence 30, Appl
30	104	6.1	859	12	US-10-369-493-25	Sequence 25, Appl
31	103.5	6.1	245	9	US-09-350-874-4	Sequence 4, Appli
32	103.5	6.1	245	15	US-10-106-989-4	Sequence 4, Appli
33	103	6.0	1038	12	US-10-032-585-7776	Sequence 7776, Ap
34	102.5	6.0	677	9	US-09-745-763-168	Sequence 168, App
35	102.5	6.0	975	12	US-10-080-608A-19	Sequence 19, Appl
36	102.5	6.0	975	12	US-10-370-685-108	Sequence 108, App
37	102.5	6.0	1106	12	US-10-032-585-7365	Sequence 7365, Ap
38	102.5	6.0	1416	12	US-10-295-027-446	Sequence 446, App
39	102	6.0	1805	11	US-09-820-843A-73	Sequence 73, Appl
40	102	6.0	1837	12	US-10-369-493-22734	Sequence 22734, A
41	102	6.0	1957	12	US-10-369-493-2070	Sequence 2070, Ap
42	101.5	6.0	1388	15	US-10-146-473-82	Sequence 82, Appl
43	101.5	6.0	1881	12	US-10-032-585-7646	Sequence 7646, Ap
44	101	5.9	1338	9	US-09-402-100-4	Sequence 4, Appli
45	100.5	5.9	1230	12	US-10-205-219-44	Sequence 44, Appl

#### ALIGNMENTS

RESULT 1

US-10-025-730-1

; Sequence 1, Application US/10025730

; Publication No. US20030045466A1

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom





```

; CURRENT FILING DATE: 2002-09-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Gal4-ANIC-BP-1
; OTHER INFORMATION: fusion protein
US-10-239-079-5

```

```

Query Match          81.0%; Score 1381; DB 12; Length 496;
Best Local Similarity 81.0%; Pred. No. 1.6e-117;
Matches 273; Conservative 31; Mismatches 29; Indels 4; Gaps 2;

```

```

Qy      4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
      || | ||||:|:| ||::|:| || | :|:|:|:|:| ||||| |||||
Db     156 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 215

Qy     60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTEVEYI 119
      || ||||| |||||:| || | :|:| ||||| ||||| |||||:| |||||
Db     216 EPQTEAVAQLAQELYNSSLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTEVEYI 275

Qy    120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
      :| ||||| |||||:| || | ||||| ||||| |||||:| || | |||||
Db    276 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 335

Qy    180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
      ||||| |||||:| || | ||||| ||||| |||||:| || | |||||
Db    336 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLDR 395

Qy    240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQP 299
      || | ||||| ||||| ||||| ||||| |||||:| || | |||||
Db    396 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVVFVANPNKTQPILDILLKNQA 455

Qy    300 KLIEFLSSPQKERTDDEQFADEKKNYLIKQIRDLKKTA 336
      ||||| || :|:| || | || | ||:| |||||:|
Db    456 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRDLKRPA 492

```

### RESULT 3

```

US-10-239-079-6
; Sequence 6, Application US/10239079
; Publication No. US20030148446A1
; GENERAL INFORMATION:
; APPLICANT: Merck Patent GmbH
; TITLE OF INVENTION: ANIC-BP1-ligand
; FILE REFERENCE: ANIC-BP-1-ligand
; CURRENT APPLICATION NUMBER: US/10/239,079
; CURRENT FILING DATE: 2002-09-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: LexA-ANIC-BP-1  
; OTHER INFORMATION: fusion protein  
US-10-239-079-6

Query Match 81.0%; Score 1381; DB 12; Length 552;  
Best Local Similarity 81.0%; Pred. No. 1.9e-117;  
Matches 273; Conservative 31; Mismatches 29; Indels 4; Gaps 2;

Qy 4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59  
|| | |||:|:| ||:|:| || | :|:|:|:| | |||| | ||||  
Db 212 MPFPFGKSHKSPADIVKNLKE\$MAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 271  
  
Qy 60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119  
|| |||||:|:| ||:|:| ||||| ||||| |||||:|:| ||||  
Db 272 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPPTVEYI 331  
  
Qy 120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179  
: |||||:|:| || ||||| |||||:| || |||:|:| ||||  
Db 332 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 391  
  
Qy 180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239  
||| |||||:| :|||:| | :||| ||||| ||||| |||||:|:|  
Db 392 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 451  
  
Qy 240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQP 299  
||| ||||| ||||| ||||| |||||:|:| |||||:|:| |||||  
Db 452 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVVFVANPNKTQPILDILLKNQA 511  
  
Qy 300 KLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLLKTA 336  
||| || | :|:| || | ||:| |||||:|  
Db 512 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRDLLKRP 548

#### RESULT 4

US-10-025-730-3

; Sequence 3, Application US/10025730  
; Publication No. US20030045466A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Gorgone, Gina A.  
; TITLE OF INVENTION: CALCIUM BINDING PROTEIN  
; FILE REFERENCE: PF-0635 US  
; CURRENT APPLICATION NUMBER: US/10/025,730  
; CURRENT FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: US/09/190,965  
; PRIOR FILING DATE: 1998-11-13  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PERL Program  
; SEQ ID NO 3  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: Mus sp.  
; FEATURE: -  
; OTHER INFORMATION: g262934

US-10-025-730-3

Query Match 80.8%; Score 1376; DB 15; Length 341;  
Best Local Similarity 80.7%; Pred. No. 2.8e-117;  
Matches 272; Conservative 32; Mismatches 29; Indels 4; Gaps 2;

```
Qy      4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
      || | |||:|:| ||::|:| || | :|:| |||:| ||||| |||||
Db      1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy     60 EPPTTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
      || ||||| |||||:|:| ||:| ||||| ||||| ||||| |||||
Db     61 EPQTEAVAQLAQELYNSGLLGTIVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy    120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
      :||| |||||:|:| || ||||| ||||| |||||:| || |||:|:| |||||
Db    121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy    180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENVYVTKRQSLKLLGELILDR 239
      ||||| ||||| |||||:| :|||:| || :||| ||||| ||||| |||||:| |||
Db    181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENVYVTKRQSLKLLGELLDR 240

Qy    240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVSPHKTQPIVEILLKNQP 299
      ||| ||||| ||||| ||||| ||||| ||||| |||||:|:| |||||:| |||||
Db    241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVVFVANPNKTQPILDILLKNQT 300

Qy    300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
      ||||| || :||:| || ||:| |||:|:|
Db    301 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRNLKRAA 337
```

RESULT 5

US-10-025-730-4

```
; Sequence 4, Application US/10025730
; Publication No. US20030045466A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gina A.
; TITLE OF INVENTION: CALCIUM BINDING PROTEIN
; FILE REFERENCE: PF-0635 US
; CURRENT APPLICATION NUMBER: US/10/025,730
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/190,965
; PRIOR FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE: -
; OTHER INFORMATION: g1794137
US-10-025-730-4
```

Query Match 65.1%; Score 1109; DB 15; Length 339;

Best Local Similarity 65.0%; Pred. No. 6.5e-93;  
Matches 217; Conservative 59; Mismatches 54; Indels 4; Gaps 3;

```
Qy      4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
      ||| || |:| |:| ||: : || |:| :|| |:|||:| ::| :| |:: |||
Db      1 MPLFGKSQKSPVELVKSLKEAINALEAGDRKVEKAQEDVSKNLVSIKNMLHGSSDAEPPA 60

Qy     64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
      : |||:|:|:| |:| || :| ||||| | | |||:|:|:|:|:|:|:|:|:|:|:|:|
Db     61 DYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNILRRQIGTRSPTVEYICTK 120

Qy    123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
      | ||| |: ||| |:|| | |||| |:| ||||:| |:| | |:|:|:|:|:|:|
Db    121 PEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180

Qy    181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLQSENYVTKRQSLKLLGELILDR 239
      |||:|:|:|:|:|:| |:|:| || | : |:| | ||||:|:|:|:|:|:|:|:|:|
Db    181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLNSENIVTRQSLKLLGELLDDR 240

Qy    240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQP 299
      ||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVVFVANPNKPKPILDILLRNQT 300

Qy    300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLLK 333
      |:|:|:|:| :|:|:| || | ||||:|:|
Db    301 KLVDFLTNFHTDRSEDEQFNDEKAYLIKQIKELK 334
```

#### RESULT 6

US-10-025-730-5

```
; Sequence 5, Application US/10025730
; Publication No. US20030045466A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gina A.
; TITLE OF INVENTION: CALCIUM BINDING PROTEIN
; FILE REFERENCE: PF-0635 US
; CURRENT APPLICATION NUMBER: US/10/025,730
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/190,965
; PRIOR FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE: -
; OTHER INFORMATION: g1255838
```

US-10-025-730-5

Query Match 62.4%; Score 1063.5; DB 15; Length 377;  
Best Local Similarity 60.5%; Pred. No. 1.1e-88;  
Matches 211; Conservative 53; Mismatches 68; Indels 17; Gaps 3;

Qy 4 MP-LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAM 49  
 || || ||||:|:| |:: | ||: | || | | :|||:|: |  
 Db 1 MPLLF GKSHKSPADVVKTLREVLTI LDKLPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60

Qy 50 KEILCGTNEKEPTE--AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR 106  
 | : | : || : | |||||:|: : | || | : || |||| ||||:|  
 Db 61 KSFIYGNDSEAEPSSSEHVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVQIFNNLLR 120

Qy 107 RQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFR 166  
 |||||:| | || |::| | || | :||| |||: ||||:|: |  
 Db 121 RQIGTRSPTVEYLGARPEILIQLVQGYSPDIALTCGLMLRESIRHDHLAKIILYSDVIFY 180

Qy 167 DFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKR 226  
 || ||: |||:|||:|:| |||| :|:|: |||| | |: || |:|:|:|  
 Db 181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAI IAEFLDSNYDTFFAQYQNLNSKNYVTRR 240

Qy 227 QSLKLLGELILDRHNFAMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFASPHK 286  
 |||||:|:| |||| | |||| :|:| |||| | ||: |||||:|:|  
 Db 241 QSLKLLGELLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAHFVFKVVFANPNK 300

Qy 287 TQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKKT 335  
 :|| :|| :|: ||:| || | :||| || | ||||:|:| :  
 Db 301 PKPISDILNRNREKLVEFLSEPHNDRTDDEQFNDEKAYLIKQIQEMKSS 349

RESULT 7

US-10-029-386-32324  
 ; Sequence 32324, Application US/10029386  
 ; Publication No. US20030194704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharron G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES  
 USEFUL FOR GENE  
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 ; FILE REFERENCE: AEOMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 34288  
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 32324  
 ; LENGTH: 820  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC000066.1  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.87  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3  
 ; OTHER INFORMATION: SWISSPROT HIT: Q99996, EVALUATE 0.00e+00  
 US-10-029-386-32324

Query Match

7.5%; Score 128.5; DB 12; Length 820;

Best Local Similarity 20.1%; Pred. No. 0.0072;  
Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;

```
Qy      18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
      :: ||||| | || : | |:: :: |: ||: || |
Db      358 IEKLKDNLGIIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE----- 404

Qy      78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI 125
      :: : ||| |:: : :: || | |::| || |:: : :
Db      405 --ISKLDLQQLVNSKSEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 460

Qy      126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
      | :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      461 LEKQMKEKE-----NDLQEKFAQLEAEN-SILKDEKK 491

Qy      186 TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
      | :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      492 TLEDMLKIHTPVSQEERLIFLDSIKSKSDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 551

Qy      238 DRHNFAIMTK-----YISKPENLKLMMNLLRD 264
      | :|: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      552 QRNTFSFAEKNFEVNYQELQEEYACLLKVDDLEDSKNKQLEYKSKLKALNEELHLQRI 611

Qy      265 KSPNIQFEA--FHVFKVVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
      :: :: | | || | :| : |:: |: :|:| | :|:| : : : |
Db      612 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMKLEVTKREKLELSQRLSDL 668

Qy      321 -----EKNYLIKQIRDLKK 334
      | :| |::: ||:
Db      669 SEQLKQKHGEISFLNEEVKSLKQ 691
```

#### RESULT 8

US-10-080-608A-11

; Sequence 11, Application US/10080608A  
; Publication No. US20030198956A1  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/10/080,608A  
; CURRENT FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 3878  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-080-608A-11

Query Match 7.5%; Score 128.5; DB 12; Length 3878;  
Best Local Similarity 20.1%; Pred. No. 0.065;  
Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;

```
Qy      18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
```

```

      :: ||||| | || : | |:: :: |: ||: || |
Db      664 IEKLDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE----- 710
Qy      78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI 125
      :: : ||| |:: : :: || | |::| || |:: :
Db      711 --ISKLDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 766
Qy      126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
      | :| | | :| :| | :|
Db      767 LEQQMKEKE-----NDLQEKFAQLEAEN-SILKDEKK 797
Qy      186 TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
      | :|:| | :| :| :| :| :| || | :| :| | :
Db      798 TLEDMLKIHTPVSQEERLIFLDSIKSKSDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 857
Qy      238 DRHNFAIMTK-----YISKPENLKLMMNLLRD 264
      |: | : | || : | :| |
Db      858 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLED SKNKQELEYSKLKALNEELHLQRI 917
Qy      265 KSPNIQFEA--FHVFKV FVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
      :: :: | | || :| : |:: |: :|:| | :|: : :: :|
Db      918 NPTTVKMKSSVFEDEKTFVA---ETLEMGEVVEKDTTELMKLEVTKREKLELSQRLSDL 974
Qy      321 -----EKNYLIKQIRDLKK 334
      | :| :::: ||:
Db      975 SEQLKQKHGEISFLNEEVKSLKQ 997

```

RESULT 9

US-10-171-311-4

; Sequence 4, Application US/10171311

; Publication No. US20030087270A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Chen, Yan

; APPLICANT: Zhao, Xumei

; APPLICANT: Monahan, John

; APPLICANT: Kamatkar, Shubhangi

; APPLICANT: Glatt, Karen

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Hoersh, Sebastian

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY

; TITLE OF INVENTION: OF CERVICAL CANCER

; FILE REFERENCE: MRI-035

; CURRENT APPLICATION NUMBER: US/10/171,311

; CURRENT FILING DATE: 2002-06-12

; PRIOR APPLICATION NUMBER: US 60/298,159

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/298,155

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/335,936

; PRIOR FILING DATE: 2001-11-14

; NUMBER OF SEQ ID NOS: 238

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 3899



; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-171-311-4

Query Match 7.5%; Score 128.5; DB 15; Length 3899;  
Best Local Similarity 20.1%; Pred. No. 0.066;  
Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;

```
Qy      18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTTEAVAQLAQELYSSG 77
      :: ||||| | || : | :|: :: |: ||: || |
Db      652 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE----- 698

Qy      78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI 125
      :: : ||| |:: : :: || | |::| || | : : :
Db      699 --ISKLKDLQQLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 754

Qy      126 LFMLLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
      | :| | | :| :| :| :| :| :| :| :| :| :|
Db      755 LEKQMKEKE-----NDLQEKFAQLEAEN-SILKDEKK 785

Qy      186 TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
      | :|:| | :| :| :| :| :| :| :| :| :| :| :|
Db      786 TLEDMLKIHTPVSQEERLIFLDSIKSKSDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 845

Qy      238 DRHNFAIMTK-----YISKPENLKLMMNLLRD 264
      | :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      846 QRNTFSFAEKNFEVNYQELQEYACLKVKDDLEDSEKSKQLELEYKSKLKALNEELHLQRI 905

Qy      265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
      :: :: | | ||| :| :| :| :| :| :| :| :| :| :| :|
Db      906 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMKLEVTKREKLELSQRLSDL 962

Qy      321 -----EKNYLIKQIRDLKK 334
      | :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      963 SEQLKQKHGEISFLNEEVKSLKQ 985
```

#### RESULT 10

US-10-171-311-2

; Sequence 2, Application US/10171311

; Publication No. US20030087270A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Chen, Yan

; APPLICANT: Zhao, Xumei

; APPLICANT: Monahan, John

; APPLICANT: Kamatkar, Shubhangi

; APPLICANT: Glatt, Karen

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Hoersh, Sebastian

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY

; TITLE OF INVENTION: OF CERVICAL CANCER

; FILE REFERENCE: MRI-035

; CURRENT APPLICATION NUMBER: US/10/171,311

; CURRENT FILING DATE: 2002-06-12

; PRIOR APPLICATION NUMBER: US 60/298,159



```
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 100
; LENGTH: 3911
; TYPE: PRT
; ORGANISM: human
US-10-370-685-100
```

```
Query Match          7.5%; Score 128.5; DB 12; Length 3911;
Best Local Similarity 20.1%; Pred. No. 0.066;
Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;
```

```
Qy      18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTAEVAQLAQELYSSG 77
      :: ||||| | || : | |:: :: |: |:: || |
Db      664 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE----- 710

Qy      78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI 125
      :: : ||| |:: : :: || | |::| | |:: |:: :
Db      711 --ISKLDLQQLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 766

Qy      126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
      | :| | | |:: | :| | |
Db      767 LEKQMKEKE-----NDLQEKFAQLEAEN-SILKDEKK 797

Qy      186 TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
      | :|:| | ::: | : : : |::| :| || | :| :| |
Db      798 TLEDMLKIHTPVSQEERLIFLDSIKSKSDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 857

Qy      238 DRHNFAIMTK-----YISKPENLKLMMNLLRD 264
      |: |: | | || : | :| |
Db      858 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 917

Qy      265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
      :: :: | | || | :| : |:: |: :|:| | :|: : : :
Db      918 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMKLEVTKREKLELSQRLSDL 974

Qy      321 -----EKNYLIKQIRDLKK 334
      | :| :::: ||:
Db      975 SEQLKQKHGEISFLNEEVKSLKQ 997
```

# RESULT 12

US-10-171-311-8

; Sequence 8, Application US/10171311

; Publication No. US20030087270A1

## GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Chen, Yan

; APPLICANT: Zhao, Xumei

; APPLICANT: Monahan, John

; APPLICANT: Kamatkar, Shubhangi

; APPLICANT: Glatt, Karen

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Hoersh, Sebastian

```
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 3917
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-8
```

```
Query Match          7.5%; Score 128.5; DB 15; Length 3917;
Best Local Similarity 20.1%; Pred. No. 0.066;
Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;
```

```
Qy      18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTTEAVAQLAQELYSSG 77
      :: ||||| | || : | |: :: |: ||: || |
Db      652 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE----- 698

Qy      78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI 125
      :: : ||| |: : :: || | |: :| || |: : :
Db      699 --ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 754

Qy      126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
      | :| | | | | | | | | | | | | | | | | | | | | |
Db      755 LEQMKEKE-----NDLQEKFAQLEAEN-SILKDEKK 785

Qy      186 TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
      | :|:| | : :: | : : : |: :| : :| || |:| :| | :
Db      786 TLEDMLKIHTPVSQEERLIFLDSIKSKSDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 845

Qy      238 DRHNFAIMTK-----YISKPENLKLMMNLLRD 264
      |: |: | | | | | | | | | | | | | | | | | | |
Db      846 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNQELEYKSKLKALNEELHLQRI 905

Qy      265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
      :: :: | | ||| :| : |: |: |: |:| | :|: : :|
Db      906 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMKLEVTKREKLELSQRLSDL 962

Qy      321 -----EKNYLIKQIRD LKK 334
      | :| : ::| ||:
Db      963 SEQLKQKHGEISFLNEEVKSLKQ 985
```

RESULT 13

US-10-171-311-6

; Sequence 6, Application US/10171311

; Publication No. US20030087270A1

```
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-6
```

```
Query Match          7.5%; Score 128.5; DB 15; Length 3925;
Best Local Similarity 20.1%; Pred. No. 0.067;
Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;
```

```
Qy      18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
      :: ||||| | || : | |: : : |:      ||:      || |
Db      652 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE----- 698

Qy      78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI 125
      :: : ||| |:: : : : || | |::|      ||      |: :      :
Db      699 --ISKLDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 754

Qy      126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
      | :| |      | : :| :|      |
Db      755 LEKQMKEKE-----NDLQEKFAQLEAEN-SILKDEKK 785

Qy      186 TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
      | :|:| |      : : : | : : : |::|      :| ||      |:| :| | :|
Db      786 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 845

Qy      238 DRHNFAIMTK-----YISKPENLKLMMNLLRD 264
      |: | : |      | || : | : :| |
Db      846 QRNTFSFAEKNFEVNYQELQEYACLLKVKDDLED SKNKQELEYKSKLKALNEELHLQRI 905

Qy      265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
      : : : | | ||| :| : |:: |: :|:| | :|: : : : :|
Db      906 NPTTVKMKSSVFDEDEKTFVA---ETLEMGEVVEKDTTELMKLEVTKREKLELSQRLSDL 962
```

Qy 321 -----EKNYLIKQIRDLKK 334  
| :: | :::: ||:  
Db 963 SEQLKQKHGEISFLNEEVKSLKQ 985

RESULT 14

US-09-864-761-47959

; Sequence 47959, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES  
USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

```
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47959
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AJ010770.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q99323, EVALUE 3.00e-17
; OTHER INFORMATION: EST_HUMAN HIT: AU132932.1, EVALUE 1.00e-105
US-09-864-761-47959
```

```
Query Match          7.3%; Score 125; DB 9; Length 660;
Best Local Similarity 20.5%; Pred. No. 0.011;
Matches 75; Conservative 68; Mismatches 119; Indels 104; Gaps 13;
```

```
Qy      18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
      :: ||||| | || : | |:: :: |: ||: || |
Db      342 IEKLKDNLGIHYKQ--QIDGLQNMESQKIETMQ-----FEKDNLITKQNQLILE----- 388

Qy      78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI 125
      :: : ||| |:: : :: || | |::| || |:: :
Db      389 --ISKLDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 444

Qy      126 LFMLLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
      | :| | | |::| :| :| | |
Db      445 LEQMKKEKE-----NDLQEKFAQLEAEN-SILKDEKK 475

Qy      186 TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
      | :|:| | :: | :: :: |::| ::| || |::| |:
Db      476 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 535

Qy      238 DRHNFAIMTK-----YISKPENLKLMMNLLRD 264
      |: |: | | || : | ::| |
Db      536 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 595

Qy      265 KSPNIQFEA--FHVFKVVFVSPHKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEK 322
      :: :: | | ||| :| : |:: |: :|:| | ::|: : | :
Db      596 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMKLEVTKREKLELSQRLSDL 652

Qy      323 NYLIKQ 328
      : :||
Db      653 SEQLKQ 658
```

# RESULT 15

US-10-023-634-18

```
; Sequence 18, Application US/10023634
; Publication No. US20030236389A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Ballinger, Robert A
; APPLICANT: Guo, Xiaojia
; APPLICANT: Tchernev, Velizar T
```

```

; APPLICANT: Shenoy, Suresh G
; APPLICANT: Li, Li
; APPLICANT: Ellerman, Karen
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Burgess, Catherine E
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gangolli, Esha A
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Gunther, Erik
; APPLICANT: Smithson, Glennda
; APPLICANT: Millet, Isabelle
; APPLICANT: Gerlach, Valerie
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; TITLE OF INVENTION: Using the Same
; FILE REFERENCE: 21402-221
; CURRENT APPLICATION NUMBER: US/10/023,634
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: 60/256,025
; PRIOR FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: 60/265,163
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/272,929
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/274,864
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/276,688
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/277,880
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/286,409
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/309,246
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/315,600
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-634-18

```

```

Query Match          6.8%; Score 116.5; DB 12; Length 709;
Best Local Similarity 19.6%; Pred. No. 0.073;
Matches 79; Conservative 64; Mismatches 146; Indels 115; Gaps 12;

```

```

Qy      3 KMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62
         |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: :
Db      179 KLQVTQRSLEESQKGIAQLEGKLVSIEKE--KIDEKS-ETEKLLLEYIEEISCASDQVEKY 235

Qy      63 TEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
         :||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```



Db 236 KLDIAQLEENL-----KEKNDEILSLKQSLEENIVILSKQVE----- 272  
 Qy 123 PHILFMLLKGYEAPQIALRCGIMLRECIRH-----EPLAKIILFSNQFRD 167  
                                   : :| : :| | : | : : | :  
 Db 273 -----DLNVKCQLEKEKEDHVNRNREHNENLNAEMQNLKQKFILEQQERE 318  
 Qy 168 FFKYVELSTFDIAS-----DAFATPKDLLTRHKVLVADFLEQNYDTIFEDYEKLL 217  
           : || : :| : : : | | : | : | : : : |  
 Db 319 KLQOKELQIDSLLOQEKELSSSLHQKLCSFQEMVKEKNLFEEELKQTLDELDKLQKKEE 378  
 Qy 218 QSENYV-----TKRQSLKLLGELI-----LDRHNFA-----IMTKY----- 248  
           | : | | : : : ||| | : : : | : : : | : ||  
 Db 379 QAERLVKQLEEEAKSRAEELKLLLEKLGKEAELEKSSAAHTQATLLLQEKYDSMVQSLE 438  
 Qy 249 -----ISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILL 295  
                           | : | : ||| : | : : | | : | : | : ||  
 Db 439 DVTAQFEGYKALTASEIEDLKLENSSLQEKAAKAGKNAEDVQHQILATESSNQEYVRMLL 498  
 Qy 296 KNQPK-----LIEFLSSFQKERTD-DEQFADEKKNYLIKQIRD 331  
           | | : | || : : || | : : || : |  
 Db 499 DLQTKSALKETEIKEITVSFLQKITDLQNQLKQQEEDFRKQLED 542

Search completed: January 7, 2004, 16:52:26  
 Job time : 37 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 16:44:17 ; Search time 21 Seconds  
(without alignments)  
1543.278 Million cell updates/sec

Title: US-10-088-872-2  
Perfect score: 1704  
Sequence: 1 MKKMPLFSKSHKNPAEIVKI.....FADEKNYLIKQIRDLKKTAP 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	1376	80.8	341	2	I57997	hypothetical calci
2	1063.5	62.4	377	2	T16651	hypothetical prote
3	1006.5	59.1	338	2	T27129	hypothetical prote
4	834.5	49.0	329	2	T50117	mo25 homolog [impo
5	685	40.2	305	2	G71441	hypothetical prote
6	632	37.1	348	2	B84448	hypothetical prote
7	485	28.5	399	2	S34681	hypothetical prote
8	143.5	8.4	339	2	T33477	hypothetical prote
9	134.5	7.9	677	2	H64574	DNA topoisomerase
10	128	7.5	430	2	H64709	hypothetical prote
11	125.5	7.4	298	2	B71685	hypothetical prote
12	125.5	7.4	1642	2	T08880	NMDA receptor-bind
13	123.5	7.2	1285	2	B72420	hypothetical prote

14	120	7.0	1175	2	F64489	hypothetical prote
15	118.5	7.0	959	2	T00246	DNA polymerase V -
16	115	6.7	474	2	S71322	glutathione syntha
17	113.5	6.7	833	2	T43446	hypothetical prote
18	112.5	6.6	1411	2	S55123	hypothetical prote
19	111.5	6.5	725	1	JC5016	hyaluronan recepto
20	111.5	6.5	2401	2	T28676	rhoptry protein -
21	111	6.5	2166	2	G70163	hypothetical prote
22	111	6.5	2819	2	A90551	conserved hypothet
23	109.5	6.4	457	2	C82911	hypothetical prote
24	109.5	6.4	978	2	A70387	conserved hypothet
25	109.5	6.4	1830	2	E82909	conserved hypothet
26	109	6.4	695	2	T07283	hypothetical prote
27	109	6.4	1401	2	S11527	alpha-latrotoxin p
28	108.5	6.4	442	2	T18507	hypothetical prote
29	108.5	6.4	952	2	T50451	hypothetical coile
30	108.5	6.4	1163	2	D64315	type I restriction
31	108	6.3	568	2	S73254	replication helica
32	107.5	6.3	483	2	I40055	positive trans-act
33	107.5	6.3	855	2	E90106	importin beta-1 SU
34	107.5	6.3	1042	2	G64514	type I restriction
35	107.5	6.3	1726	1	SAZQGM	major merozoite su
36	107.5	6.3	1726	2	A45948	major merozoite su
37	107	6.3	570	2	S68686	phosphoprotein pho
38	107	6.3	1173	2	T43527	sp8 protein - fiss
39	107	6.3	1727	2	T50073	myosin-like coiled
40	106	6.2	474	2	S56748	glutathione syntha
41	106	6.2	1295	2	T24587	hypothetical prote
42	105.5	6.2	781	2	T00456	protein kinase hom
43	105.5	6.2	847	2	A56039	GTPase-activating
44	105.5	6.2	1091	2	T34107	hypothetical prote
45	105.5	6.2	1619	2	T18499	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

I57997

hypothetical calcium-binding protein - mouse

C;Species: Mus sp. (mouse)

C;Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 19-May-2000

C;Accession: I57997

R;Miyamoto, H.; Matsushiro, A.; Nozaki, M.

Mol. Reprod. Dev. 34, 1-7, 1993

A;Title: Molecular cloning of a novel mRNA sequence expressed in cleavage stage mouse embryos.

A;Reference number: I57997; MUID:93119656; PMID:8418809

A;Accession: I57997

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-341 <RES>

A;Cross-references: GB:S51858; NID:g262933; PIDN:AAB24801.1; PID:g262934

C;Superfamily: Saccharomyces hypothetical protein YKL189w

C;Keywords: calcium binding

Query Match

80.8%; Score 1376; DB 2; Length 341;

Best Local Similarity 80.7%; Pred. No. 7.9e-85;  
Matches 272; Conservative 32; Mismatches 29; Indels 4; Gaps 2;

```
Qy      4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
      || | |||:|:| | |::|:| | | | :|:| | | | | | | | | | | | |
Db      1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy     60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     61 EPQTEAVAQLAQELYNSGLLGLTVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy    120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
      : | | | | | | | | | | : | | | | | | | | | | | | | | | | | | |
Db    121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy    180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
      | | | | | | | | | | | | : | | | | : | | | | | | | | | | | | | |
Db    181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLDR 240

Qy    240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300

Qy    300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKA 336
      | | | | | | | | : | | : | | | | | | | | | | | | | | | |
Db    301 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRNLKRAA 337
```

## RESULT 2

T16651

hypothetical protein R02E12.2 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000

C;Accession: T16651

R;Leimbach, D.

submitted to the EMBL Data Library, April 1996

A;Description: The sequence of *C. elegans* cosmid R02E12.

A;Reference number: Z18554

A;Accession: T16651

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-377 <LEI>

A;Cross-references: EMBL:U53337; NID:g1255833; PID:g1255838; PIDN:AAA96187.1;

GSPDB:GN00028; CESP:R02E12.2

A;Experimental source: strain Bristol N2; clone R02E12

C;Genetics:

A;Gene: CESP:R02E12.2

A;Map position: X

A;Introns: 37/3; 146/2; 225/1; 315/3

C;Superfamily: *Saccharomyces* hypothetical protein YKL189w

Query Match 62.4%; Score 1063.5; DB 2; Length 377;

Best Local Similarity 60.5%; Pred. No. 7e-64;

Matches 211; Conservative 53; Mismatches 68; Indels 17; Gaps 3;

```
Qy      4 MP-LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAM 49
      || || |||:|:| | |::| | | | | | | | | | | | | | | | | | | |
```

Db 1 MPLLF GKSHKSPADVVKTLREVLTLIDKLPKPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60

Qy 50 KEILCGTNEKEPTE---AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR 106  
 | : | : | : | | | | | : : : | | | : | | | | | | : |

Db 61 KSFIYGNDSEPSSEHVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVQGI FNNLLR 120

Qy 107 RQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFR 166  
 | | | | | | | | : | | | | : : | | | | | : | | | | | : | |

Db 121 RQIGTRSPTVEYLGARPEILIQLVQGYSPDIALTCGLMLRESIRHDHLAKIILYSDVFI 180

Qy 167 DFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKR 226  
 | | | : | | | : | | | : | | | : : | | | : | | | : | | | : |

Db 181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAI IAEFLDSNYDTFFAQYQNLNLSKNYVTRR 240

Qy 227 QSLKLLGELILDRHNFAMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHK 286  
 | | | | | | | : | | | | | | | : | | | : | | | | | | | : | | | | | : |

Db 241 QSLKLLGELLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAHVFKVVFVANPNK 300

Qy 287 TQPIVEILLKNQPKLIEFLSSSQKERTDDEQFADEKKNYLIKQIRDLKKT 335  
 : | | : | | : : | | | | | : | | | | | | | | | | : : | :

Db 301 PKPISDILNRNREKLVEFLSEFHNDRTDDEQFNDEKAYLIKQIQEMKSS 349

# RESULT 3

T27129

hypothetical protein Y53C12A.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000

C;Accession: T27129

R;Kershaw, J.; Lennard, N.

submitted to the EMBL Data Library, September 1997

A;Reference number: Z20315

A;Accession: T27129

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-338 <WIL>

A;Cross-references: EMBL:Z99277; PIDN:CAB16486.1; GSPDB:GN00020; CESP:Y53C12A.4

A;Experimental source: clone Y53C12A

C;Genetics:

A;Gene: CESP:Y53C12A.4

A;Map position: 2

A;Introns: 29/3; 103/3; 136/2; 215/1; 282/3

C;Superfamily: *Saccharomyces* hypothetical protein YKL189w

Query Match 59.1%; Score 1006.5; DB 2; Length 338;

Best Local Similarity 57.2%; Pred. No. 3.9e-60;

Matches 191; Conservative 60; Mismatches 78; Indels 5; Gaps 1;

Qy 5 PLFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAMKEILCGTNEK 59  
 | | | : | | : : | | : | : : : : : | | : | | : | : | : |

Db 4 PLFGKADKTPADVVKNLRLDALLVIDRHGTNTSERKVEKAIEETAKMLALAKTFIYGSDAN 63

Qy 60 EPPTAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119  
 | | | | | | | : | : : | | | : | | | : | | | : | | | : | | | : |

Db 64 EPNNEQVTQLAQEVYNANVLPMLIKHLHKFEFECKKDVASVFNNLLRRQIGTRSPTVEYL 123

Qy 120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179









Qy 237 LDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVSPHKTQPIVEILLK 296  
 :|| | :| ||:| :||::||| | : ||| |||:|:| | : | : || ||:  
 Db 244 MDRNSGVMVKYVSSLDNLRIMMNLREPTKNIQLEAFHIFKLFVANENKPEDIVAILVA 303

Qy 297 NQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK----KTA 336  
 |: |: : : |: :| | :| :: :| | |||  
 Db 304 NRTKILRLFADLKPEK-EDVGFETDKALVMNEIATLSLLDIKTA 346

RESULT 7

S34681

hypothetical protein YKL189w - yeast (*Saccharomyces cerevisiae*)

C;Species: *Saccharomyces cerevisiae*

C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-Apr-2002

C;Accession: S34681; S33963; S38021; S38026

R;Wieman, S.; Voss, H.; Schwagaer, C.; Rupp, T.; Stegemann, J.; Zimmermann, J.; Grothues, D.; Sensen, C.; Erfle, H.; Hewitt, N.; Banrevi, A.; Ansorge, W. submitted to the EMBL Data Library, July 1993

A;Description: Sequencing and analysis of 51.5 kilobases on the left arm of chromosome XI from *Saccharomyces cerevisiae* reveals 23 open reading frames including the FAS1 gene.

A;Reference number: S34679

A;Accession: S34681

A;Molecule type: DNA

A;Residues: 1-399 <WIE>

A;Cross-references: EMBL:X74151; NID:g450365; PIDN:CAA52249.1; PID:g395236

A;Experimental source: strain S288C

R;Cheret, G.; Mattheakis, L.C.; Sor, F.

Yeast 9, 661-667, 1993

A;Title: DNA sequence analysis of the YCN2 region of chromosome XI in *Saccharomyces cerevisiae*.

A;Reference number: S33960; MUID:93348778; PMID:8394042

A;Accession: S33963

A;Molecule type: DNA

A;Residues: 1-399 <CHE>

A;Cross-references: GB:X69765; NID:g296985; PIDN:CAA49422.1; PID:g296989

R;Wiemann, S.; Voss, H.; Schwager, C.; Rupp, T.; Grothues, D.; Sensen, C.; Stegemann, J.; Zimmermann, J.; Erfle, H.; Hewitt, N.; Ansorge, W.

submitted to the Protein Sequence Database, March 1994

A;Reference number: S37825

A;Accession: S38021

A;Molecule type: DNA

A;Residues: 1-399 <WI2>

A;Cross-references: EMBL:Z28189; NID:g486334; PIDN:CAA82032.1; PID:g486335; MIPS:YKL189w

A;Experimental source: strain S288C

R;Maia e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.; Guerreiro, P.; Rodrigues-Pousada, C.

submitted to the Protein Sequence Database, March 1994

A;Reference number: S38024

A;Accession: S38026

A;Molecule type: DNA

A;Residues: 1-399 <MAI>

A;Cross-references: EMBL:Z28189; NID:g486334; PIDN:CAA82032.1; PID:g486335; MIPS:YKL189w

A;Experimental source: strain S288C

C;Genetics:  
A;Gene: SGD:HYM1  
A;Cross-references: SGD:S0001672  
A;Map position: 11L  
C;Superfamily: Saccharomyces hypothetical protein YKL189w

Query Match 28.5%; Score 485; DB 2; Length 399;  
Best Local Similarity 33.0%; Pred. No. 3.6e-25;  
Matches 113; Conservative 75; Mismatches 138; Indels 16; Gaps 6;

```
Qy      7 FSKSHKNPAEIVKILKDNLAILEK----QDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62
      : | : | | : : : |          || | | | | : | | | : | : |
Db     16 WKKNPKTPSDYARLIEQLNKFSSPSLTQDNKR-KVQEECTKYLIGTKHFIVGDTDPHPT 74

Qy     63 TEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
      || : : | : : : | : : || : : | | : | : | : | : :
Db     75 PEAIDELYTAMHRADVFEYELLHFVDLEFEARRECMLIFSICLGYSKDNKFVTVDYLVLSQ 134

Qy    123 PHILFMLLKGYE-----APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELS 175
      | : : : | | | | | | | | | | : | | | | | | : : |
Db    135 PKTISLMLRTAEVALQQKGCQDIFLTVGNMII ECI KYEQLCRIILKDPQLWKFFFEFAKLG 194

Qy    176 TFDIASDAFATFKDLLTRHKVLVA-DFL--EQNYDTIFEDYEKLLQSENYVTKRQSLKLL 232
      | : : : : | | | | : : | | : | : | : | | | | |
Db    195 NFEISTESLQILSAAFTAHPKLVSKFEFFSNEINIIRFIKCINKLMAHGSYVTKRQSTKLL 254

Qy    233 GELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVSPHKTQPIVE 292
      || : | | | : | | | | | | : | | | | | | | | | | : : :
Db    255 ASLIVIRSNNALMNIYINSPENLKLIMTLMTDKSKNLQLEAFNVFKVMVANPRKSKPVFD 314

Qy    293 ILLKNQPKLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLKK 334
      || : | : | : : : | : : | | | : : : : | | :
Db    315 ILVKNRDKLLTYFKTFGLD-SQDSTFLDEREFIVQEIDSLPR 355
```

#### RESULT 8

T33477

hypothetical protein T27C10.3 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C;Accession: T33477

R;Zhu, H.J.; Graves, T.; Hawkins, M.

submitted to the EMBL Data Library, October 1998

A;Description: The sequence of *C. elegans* cosmid T27C10.

A;Reference number: Z21354

A;Accession: T33477

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-339 <ZHU>

A;Cross-references: EMBL:AF098504; PIDN:AAC67411.1; GSPDB:GN00019; CESP:T27C10.3

A;Experimental source: strain Bristol N2; clone T27C10

C;Genetics:

A;Gene: CESP:T27C10.3

A;Map position: 1

A;Introns: 72/3; 120/3; 233/3; 295/1

Query Match 8.4%; Score 143.5; DB 2; Length 339;

Best Local Similarity 19.3%; Pred. No. 0.02;  
Matches 38; Conservative 50; Mismatches 76; Indels 33; Gaps 4;

```
Qy      159 ILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQ 218
          :: :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      100 LMNTNKFRD-----FDVIQGTFTLQIIFFTNHESANNFIKNNLPRFMQTLHKLIA 150

Qy      219 SENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFK 278
          | : : : | | | | | | | | : : : : : | : | : : : : : | : :
Db      151 CSNFFIQAKSFKFLNELFTAQTNYETRSLWMAEPAFIKLVVLAIQSNKHAVRSRAVSILE 210

Qy      279 VFVASPHKTQPIVEILLKNQPKLIEFL-----SSFQKERTDDEQFAD----- 320
          :|: :| : : | : : | | | | | | | | | | | | | | | | | | |
Db      211 IFIRNPRNSPEVHEFIGRNRNVLIAFFNSAPIHYQGSFNEKE---DAQYARMAYKLLN 267

Qy      321 ---EKNYLIKQIRDLKK 334
          : : : : | : | : :
Db      268 WDMQRPFTQEQLQDFEE 284
```

#### RESULT 9

H64574

DNA topoisomerase I - *Helicobacter pylori* (strain 26695)

C;Species: *Helicobacter pylori*

C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999

C;Accession: H64574

R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.;  
Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.;  
Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;  
Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzgerald,  
L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback,  
T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.;  
Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.;  
Fraser, C.M.; Venter, J.C.

A;Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.

A;Reference number: A64520; MUID:97394467; PMID:9252185

A;Accession: H64574

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-677 <TOM>

A;Cross-references: GB:AE000559; GB:AE000511; NID:g2313536; PIDN:AAD07502.1;

PID:g2313542; TIGR:HP0440

C;Superfamily: DNA topoisomerase I

Query Match 7.9%; Score 134.5; DB 2; Length 677;

Best Local Similarity 21.6%; Pred. No. 0.19;

Matches 88; Conservative 58; Mismatches 134; Indels 127; Gaps 16;

```
Qy      7 FSKSHKNPA-EIVKILKDNL-----AILEKQDKK---TDKASEEVSKSLQAMKE 51
          | | | : : | | | | | | | | : | : | | | : | | | |
Db      222 FKFKDKNEASQFLKDLKDGLGMSVLVSLKESLSNKEPKKPFTTSKLLSQASKSLKI --- 278

Qy      52 ILCGTNEKEPPTEAVAQIAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGT 111
          || : ||| | | : : | | : | : | | | | |
```

```

Db      279 -----PTKEIAQLAQLKFEAGLITYHRTDSEFLSPEYLKEHEVFFFEPIY----- 322
Qy      112 RSPTV----EYIS-----AHPHILFMLLKGYEAPQIALRCGIMLRECIRHE 153
      | : |      | | :      | | | | | | : | : : | : |
Db      323 --PSVYQYREYKAGKNSQAEAHEAIRITHPHALKDLEKVCSDAKISEELALKLYQLIYTN 380
Qy      154 PL---AKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIF 210
      : : : | : | | : | |      | | : |      | | : | : | : |
Db      381 TICSQSRNALY-NQYDCIFK-----IKSESFKLSFKLLKEKGFLEIEELIQGKEEIN 431
Qy      211 EDYEKLLQSENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQ 270
      : | : : | | : |      | | : : : : | : | | : | : | : |
Db      432 RE-EQESEIENFSLKENDSVPLKEVFIKK-----IEKPSPKPYKESAFIPLLESEG---- 481
Qy      271 FEAFHVFKVVFASPHKTQPIVEILLKNQ-----PKLIEFLSSFQKERTDD- 315
      : |      : : | | :      : : | : | | : | : | : |
Db      482 -----IGRPSTYASFLDLLLKRKYISIDTKTNAITPTSQGLEVISFFKKDKKEVDF 531
Qy      316 -----EQF-----ADEKNYLIKQIRDLKKTA 336
      : |      | : : : : | | | |
Db      532 IALTSKDKSKLGNNTTKQFEECLDLIMRGEASYEKFMLEVISKLKSTA 578

```

# RESULT 10

H64709

hypothetical protein HP1520 - *Helicobacter pylori* (strain 26695)

C;Species: *Helicobacter pylori*

C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999

C;Accession: H64709

R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M.; Venter, J.C.

A;Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.

A;Reference number: A64520; MUID:97394467; PMID:9252185

A;Accession: H64709

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-430 <TOM>

A;Cross-references: GB:AE000650; GB:AE000511; NID:g2314700; PIDN:AAD08565.1; PID:g2314705; TIGR:HP1520

C;Superfamily: *Helicobacter pylori* hypothetical protein HP1520

Query Match 7.5%; Score 128; DB 2; Length 430;  
 Best Local Similarity 20.9%; Pred. No. 0.29;  
 Matches 82; Conservative 73; Mismatches 135; Indels 102; Gaps 20;

```

Qy      7 FSKSHKNPAEI----VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62
      | : | : | |      | | | | : | : | : | : | : | : |
Db      60 FYPNRKSKIEIEFNGEKILKENVAVFHSYDE--EFSSDSVTTFMAKSDL-----KQQY 111

```

Qy 63 TEAVAQLAQELYSSGLLVTL--IA-----DLQLIDFEGKKDVTQIFNNILR----- 106  
 : : | : | | : | | : : | | : | : |  
 Db 112 DNILLELEKE--KKALLKSLRDIASGFDYEEIEIKTIKNEKNKSFYEILDNHLTEIESSEK 169

Qy 107 -----RQIGTRSPTV-EYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKII 159  
 | | | : : : | : : | :  
 Db 170 HYSFKYRDIFDGSKKVKDFVNKHHDLEQYFNKYQ-----ELLSQSK 211

Qy 160 LF-----SNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQ----- 204  
 : | : | | : | : | : : : : | : | : |  
 Db 212 IFKHMNSGDFGTNHADDLKKALENNRFFKANHSLKIAGEEITNYQKL-SDIFENEKNRIL 270

Qy 205 NYDTIFEDYEKLLQSENYVTKRQSLKLLGELI-----LDRHNF--AIMTKYISKP 252  
 | : : | : : | : : | : | : | : : | : :  
 Db 271 NNEELKESFDKI---EKVINANKELKAFKDAISKDNTLLTEFLDYDSFRKKVLFPSYLKQV 327

Qy 253 -ENLKLMMNLLRDKSPNIQFEAFHVFKVVFASPHKTQPIVEILLKNQPKLIEFLSSFQKE 311  
 : : | : : | | : | : : : : | | : | : :  
 Db 328 IQNVKSLVNLVREKKPEIE---EIIKQASKDQKEWESVIEIF--NQRFLVPFKVELQNQ 381

Qy 312 R-----TDDEQ----FADEKNYLIKQIRDLKK 334  
 : | | : : : | | : |  
 Db 382 KDILLNKDAAQFRFIFSDDNQDMNVQKEDLQK 413

# RESULT 11

B71685

hypothetical protein RP295 - Rickettsia prowazekii

C;Species: Rickettsia prowazekii

C;Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000

C;Accession: B71685

R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland, C.G.

Nature 396, 133-140, 1998

A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A;Reference number: A71630; MUID:99039499; PMID:9823893

A;Accession: B71685

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-298 <AND>

A;Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14756.1;

PID:g3860856; GSPDB:GN00081

A;Experimental source: strain Madrid E

C;Genetics:

A;Gene: RP295

Query Match 7.4%; Score 125.5; DB 2; Length 298;  
 Best Local Similarity 20.1%; Pred. No. 0.27;  
 Matches 62; Conservative 57; Mismatches 114; Indels 75; Gaps 13;

Qy 73 LYSSGLLVTLIADLQLIDFEGKKDVTQ-----IFNNILRRQIGTRS 113  
 | : | : | : : : : | : | : : | : |  
 Db 6 LFIQLLVITSLVKAETIEVDSLNKITQDFKVNYNKNYLPQDLLVVTVLDDKFLFKSFGV-- 63

Qy	114	PTVEYISAHPHILFMLLKGY--EAPQIALRCGIMLRECIRHEPLAKIILFSNQFR-----	166
		: :   : :   :   : :     :   :	
Db	64	PIGEYIDQHRYLALAPLFSHINKNPKIIY-----ITQLILTNNSYKKELQE	109
Qy	167	-DFFKYV-ELSTFDI----ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSE	220
		:     :   : :   : : : :     :   : :   :   :	
Db	110	SDFPNFVNEMSNSQIP IIAVNGGFTGNFN NIPKF EIW FADYL KKNF---YIDFSKSFPNN	166
Qy	221	NYVTKRQSLKLLGELILDRHNFAIMTKYISKPENL--KLMMNLLRDKSPNIQFEAFHVF	277
		: :   : : :       :   :   :	
Db	167	NYI-----IFNNLDSFDNTYPVFYKGILTSNNIPASKVILNFL-----IQINFIPKC	213
Qy	278	KVFVASPHKTQPIVEILLKNQPKLIEFLSSF--QKERTDDEQFADEKNY-----LIKQI	329
		: :   : :         :   :   :	
Db	214	FILISSRELLRSMEFQLNNYSSNILFIGYHYNNKSISDDKDYKDIAYYTKMINDLIPQI	273
Qy	330	RDLKKTAP 337	
		:	
Db	274	NKIKRN NP 281	

Query Match 7.4%; Score 125.5; DB 2; Length 1642;  
Best Local Similarity 20.2%; Pred. No. 2.4;  
Matches 77; Conservative 73; Mismatches 117; Indels 115; Gaps 15;



```

      :::|   ||| |:   ::|::| : :|   : |||:: | : | :|
Db      671 GITYEILDGVVYFEGTKENVEKAKELFSDIVEK-LGEVRKEETVEFLEVNSSFPVDEFIN 729

Qy      130 LKGYEAPQIALRCGIMLRECIRHEPLAKIIL-----FSNQFRDFF----KYVELST 176
      | |   | :   | : | ::|   | : :| ||   | | : :
Db      730 LSGKLYPDVT-----CFSLDQLGLLVLKGSSEAVEDLSSMYRSFFERHQKIVKENV 780

Qy      177 FD---IASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLG 233
      ||   :   : :|:: | :||| : :   ::| ||   | |:: : :| |
Db      781 FDRIMLEVPSGFSFEFEKTFLEVLVPEVKQ----VVYLDKLNLLLVEVPVVSQSERVKSLL 836

Qy      234 ELILDRHNFAIMTKYIS-----KPENL-KLMNLLRDKSPNIQFEAF-HVFKV FVAS 283
      : | :   | :   | : | :   |||   | : : | : : : |
Db      837 DTFLKKEEAVSEKKAVKSVTIPSGVNPDELSSYLKKLLR----NVEITVFPNMGQMIVEG 892

Qy      284 P-HKTQPIVEILLKNQPKLIEFLSSSFQKERTDDEQFADEK 322
      | : : | : : : | : :   ||| | : : |
Db      893 PENEVEKAVELVEAEKEKIV-----LKERKDYVKVSDGK 926

```

# RESULT 14

F64489

hypothetical protein MJ1519 - Methanococcus jannaschii

C;Species: Methanococcus jannaschii

C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000

C;Accession: F64489

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.

A;Reference number: A64300; MUID:96337999; PMID:8688087

A;Accession: F64489

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1175 <BUL>

A;Cross-references: GB:U67593; GB:L77117; NID:g2826427; PIDN:AAB99538.1;

PID:g1500409; TIGR:MJ1519

C;Genetics:

A;Map position: FOR1494096-1497623

Query Match 7.0%; Score 120; DB 2; Length 1175;

Best Local Similarity 21.5%; Pred. No. 3.6;

Matches 76; Conservative 58; Mismatches 131; Indels 88; Gaps 15;

```

QY      7 FSKSHKNPAEIVKILKD-NLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 65
      |:| : : | | | |:|   || |:| :| :| : :   | : | |
Db      232 FNKFREENQDFDKYLTDENIAFRPHVMKKFDEFAENIKKVIAELE----GSKYKYPGLPG 287

QY      66 VAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSP TVEYISAHPHI 125

```



```

      |  |  | :  :  |  |  |  :  :  :  |  :  :  |  :  :
Db      288 V-----LYFLGMEDAYSRYIELWKNEGEKGEEKLYNALI-ESLENRKENLEF----- 333

Qy      126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFK----YVELSTFDIA- 180
      |  :  :  |  |  :  |  |  |  |  |  |  |  :
Db      334 -----GITKKVIDKFIAQKEEFREFLKNYAVYYELSAFKLEK 370

Qy      181 -----SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSL----- 229
      |  :  :  :  |  |  |  |  |  |  :  :  :  |  :  :  |
Db      371 IKEQYEKEFINLDNIIKNPYILVED-LKEN-----DSFERIIFEELDSWERRRLGDKFNP 424

Qy      230 -----KLLGELILDRH----NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAF---- 274
      |  |  |  |  |  |  |  |  |  |  |  |  :  |  :  :  |  :  :  |
Db      425 YSPYRVRALLVE-ILKRHLSSGNTTISTK-----DLKDFFEKMDKDIVKITFDEFLRII 477

Qy      275 HVFKVVFVASPHKTQPIVEILLKNQPKLIEFLSSFKERTDDEQFADEKKNYLIK 327
      :  |  :  :  |  :  :  :  |  :  :  :  |  :  :  :  |  :  :  |
Db      478 EEYKDIIS--EKVEIVKKEVKNNENKEIIEFLTKEIREYEEIIENTINYLLK 528

```

# RESULT 15

T00246

DNA polymerase V - fission yeast (*Schizosaccharomyces pombe*)

C;Species: *Schizosaccharomyces pombe*

C;Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 31-Jan-2000

C;Accession: T00246; T39442

R;Shimizu, K.

submitted to the EMBL Data Library, March 1998

A;Description: *S.pombe* homolog of *S.cerevisiae* DNA polymerase V.

A;Reference number: Z14129

A;Accession: T00246

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-959 <SHI>

A;Cross-references: EMBL:AB012696; NID:d1224325; PIDN:BAA32046.1; PID:d1033008

R;Xiang, Z.; Aves, S.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.  
submitted to the EMBL Data Library, March 1998

A;Reference number: Z21854

A;Accession: T39442

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-959 <LYN>

A;Cross-references: EMBL:AL022305; PIDN:CAA18436.1; GSPDB:GN00067;  
SPDB:SPBC14C8.14c

A;Experimental source: strain 972h-; cosmid c14C8

C;Genetics:

A;Gene: pol5+; SPBC14C8.14c

A;Map position: 2

A;Introns: 66/3

Query Match 7.0%; Score 118.5; DB 2; Length 959;

Best Local Similarity 20.5%; Pred. No. 3.5;

Matches 80; Conservative 63; Mismatches 135; Indels 113; Gaps 19;

```

Qy      9 KSHKN-----PAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62
      |  |  |  :  :  :  |  |  |  |  :  :  :  |  |
Db      522 KSPKNNLLISMDESVEIVQKSLSVLHKVTKKIDKKAQHL-QQLNAF----- 567

```

Qy 63 TEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKD--VTQIFNNILRRQIGTRSPTVEYI- 119  
 | | | | | | | | | | : : | | : | : | | |  
 Db 568 -----QLLYSLVLLQVYAGDTSIDVLEDIDNCYSKVFNNKSKRESTSNEPTAMEIL 619

Qy 120 -----SAHPHILF-----MLLKGY----EAPQIALRC-----GIMLRECI 150  
 : | : | | | : | | | | : | :  
 Db 620 TEVMLSLLSRPSLLLRKLVDMLFTSFSEDNMRESIHLICDVLKAKESVKDSEGMFAGEEV 679

Qy 151 RHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIF 210  
 : : : | | : : : | | : : : | | :  
 Db 680 EEDAFGE---TEMDEDDFEEIDTDEIEEQSD-----WEMISNQDASDNEELERKLDKVL 730

Qy 211 EDYEKLLQ-----SENYVTKRQSLKL---LGELILDRHNFAIMTKYISKPENLKLMMNLL 262  
 | | : : : | : | | | | : : | | : |  
 Db 731 EDADAKVKDEESSEEEELMNDEQMLALDEKLAEVFRER-----KKASNKEKKKNAQ 780

Qy 263 RDKSPNIQFEAFHVFKV--FVASPHKTQ-----PIVEILLKNQPKLIE----- 303  
 | : | | : | | : : : | | | : : | | : | : |  
 Db 781 ETKQQIVQFKV---KVIDLIDNYKYKTQPNNGLGFEFLIPLLEMILKTKHKVLEEKQAV 836

Qy 304 FLSSFQKERTDDEQFADEKKNYL--IKQIRDL 332  
 | : | : : | : | | : : : |  
 Db 837 FRNRLSKLKWTEEK-PSSKNVLEALKKVHVL 866

Search completed: January 7, 2004, 16:46:10  
 Job time : 32 secs

OM protein - protein search, using sw model

Run on: January 7, 2004, 16:44:17 ; Search time 17 Seconds  
(without alignments)  
932.235 Million cell updates/sec

Title: US-10-088-872-2  
Perfect score: 1704  
Sequence: 1 MKKMPLFSKSHKNPAEIVKI.....FADEKNYLIKQIRDLKKTAP 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1685	98.9	334	1	MO2L_HUMAN	Q9h9s4 homo sapien
2	1669	97.9	334	1	MO2L_MOUSE	Q9db16 mus musculu
3	1381	81.0	341	1	MO25_HUMAN	Q9y376 homo sapien
4	1376	80.8	341	1	MO25_MOUSE	Q06138 mus musculu
5	1111	65.2	339	1	MO25_DROME	P91891 drosophila
6	1006.5	59.1	338	1	MO2M_CAEEL	O18211 caenorhabdi
7	834.5	49.0	329	1	YFV6_SCHPO	Q9p7q8 schizosacch
8	776	45.5	321	1	DE76_CHLPR	Q9xfy6 chlorella p
9	728	42.7	343	1	MO2N_ARATH	Q9fgk3 arabidopsis
10	716.5	42.0	343	1	MO2M_ARATH	Q9m0m4 arabidopsis
11	666	39.1	384	1	HYMA_EMENI	O60032 emericella
12	632	37.1	348	1	MO2L_ARATH	Q9zq77 arabidopsis
13	485	28.5	399	1	HYM1_YEAST	P32464 saccharomyc
14	143.5	8.4	339	1	MO2L_CAEEL	Q9tzm2 caenorhabdi
15	128.5	7.5	3911	1	AKA9_HUMAN	Q99996 h a-kinase
16	125.5	7.4	298	1	Y295_RICPR	Q9zdn2 rickettsia
17	118.5	7.0	959	1	DPO5_SCHPO	O60094 schizosacch

18	116.5	6.8	724	1	HMMR_HUMAN	075330	homo sapien
19	115	6.7	474	1	GSHB_MOUSE	P51855	mus musculu
20	112.5	6.6	1411	1	YM42_YEAST	Q03214	saccharomyc
21	109.5	6.4	978	1	RA50_AQUAE	O67124	aquifex aeo
22	109	6.4	695	1	YCX7_CHLVU	O20159	chlorella v
23	109	6.4	1401	1	LATA_LATMA	P23631	latrodectus
24	108.5	6.4	586	1	2A5D_RABIT	Q28653	o serine/th
25	108.5	6.4	602	1	2A5D_HUMAN	Q14738	h serine/th
26	108.5	6.4	1075	1	Y124_METJA	Q57588	methanococc
27	108	6.3	568	1	DNAB_PORPU	P51333	porphyra pu
28	107.5	6.3	483	1	ACPA_BACAN	Q44643	bacillus an
29	107.5	6.3	1042	1	T1RH_METJA	Q60295	methanococc
30	107.5	6.3	1726	1	MSP1_PLAFC	P04934	plasmodium
31	107.5	6.3	1726	1	MSP1_PLAFP	P50495	plasmodium
32	107	6.3	1727	1	ALM1_SCHPO	Q9utk5	schizosacch
33	106	6.2	474	1	GSHB_HUMAN	P48637	homo sapien
34	105.5	6.2	793	1	REGA_DICDI	Q23917	dictyosteli
35	105.5	6.2	847	1	RSG2_RAT	Q63713	rattus norv
36	104.5	6.1	1701	1	MSP1_PLAFF	P13819	plasmodium
37	104.5	6.1	1701	1	MSP1_PLAFM	P08569	plasmodium
38	104	6.1	859	1	MUTS_AQUAE	O66652	aquifex aeo
39	104	6.1	1290	1	RA50_SCHPO	Q9utj8	schizosacch
40	104	6.1	1682	1	MSP1_PLAF3	P19598	plasmodium
41	103.5	6.1	641	1	PRIM_UREPA	Q9ppz6	ureaplasma
42	103	6.0	2663	1	CENE_HUMAN	Q02224	homo sapien
43	102.5	6.0	502	1	URIC_BACSB	Q45697	bacillus sp
44	102.5	6.0	975	1	KINH_DROME	P17210	drosophila
45	102.5	6.0	1202	1	RPM2_YEAST	Q02773	saccharomyc

# ALIGNMENTS

## RESULT 1

### MO2L\_HUMAN

ID MO2L\_HUMAN STANDARD; PRT; 334 AA.  
AC Q9H9S4; Q9BZ33;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE MO25-like protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE OF 4-334 FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
RA Ninomiya K., Iwayanagi T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]

RP SEQUENCE OF 276-334 FROM N.A.  
RA Pearce A.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: Belongs to the Mo25 family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AK022639; BAB14147.1; ALT\_INIT.  
DR EMBL; AL138875; CAC28084.1; -.  
DR InterPro; IPR004892; Mo25.  
DR Pfam; PF03204; Mo25; 1.  
SQ SEQUENCE 334 AA; 38728 MW; 97702273D8548432 CRC64;

Query Match 98.9%; Score 1685; DB 1; Length 334;  
Best Local Similarity 99.7%; Pred. No. 1.3e-100;  
Matches 333; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
      |||
Db      1 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 60

Qy     64 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHP 123
      |||
Db     61 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHP 120

Qy    124 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 183
      |||
Db    121 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLVKIILFSNQFRDFFKYVELSTFDIASDA 180

Qy    184 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENVYTKRQSLKLLGELILDRHNFA 243
      |||
Db    181 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENVYTKRQSLKLLGELILDRHNFA 240

Qy    244 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQPKLIE 303
      |||
Db    241 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQPKLIE 300

Qy    304 FLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAP 337
      |||
Db    301 FLSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAP 334

```

## RESULT 2

### MO2L\_MOUSE

ID MO2L\_MOUSE STANDARD; PRT; 334 AA.  
AC Q9DB16; Q8BG52; Q91WB8; Q91YL0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE MO25-like protein.  
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=C57BL/6J;  
 RC TISSUE=Cerebellum, Eye, Pituitary, and Testis;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=FVB/N; TISSUE=Mammary gland, and Salivary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9DB16-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9DB16-2; Sequence=VSP\_007417, VSP\_007418;  
 CC Note=No experimental confirmation available;  
 CC -!- SIMILARITY: Belongs to the Mo25 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AK005323; BAB23953.2; ALT\_INIT.  
 DR EMBL; AK030474; BAC26978.1; ALT\_INIT.  
 DR EMBL; AK053642; BAC35457.1; ALT\_INIT.  
 DR EMBL; AK076758; BAC36470.1; ALT\_INIT.  
 DR EMBL; AK076867; BAC36513.1; -.  
 DR EMBL; BC016128; AAH16128.1; -.  
 DR EMBL; BC016546; AAH16546.1; -.  
 DR MGD; MGI:1916258; 1500031K13Rik.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 KW Alternative splicing.  
 FT VARSPLIC 276 293 VFVASPHKTQPIVEILLK -> NSVFITNRIHGLKRWLSS  
 FT (in isoform 2).  
 FT /FTid=VSP\_007417.  
 FT VARSPLIC 294 334 Missing (in isoform 2).  
 FT /FTid=VSP\_007418.  
 FT CONFLICT 42 42 S -> P (IN REF. 1; BAB23953).  
 FT CONFLICT 229 229 L -> R (IN REF. 2; AAH16546).  
 SQ SEQUENCE 334 AA; 38718 MW; 822F04A87FB4EB6F CRC64;

Query Match 97.9%; Score 1669; DB 1; Length 334;  
 Best Local Similarity 98.5%; Pred. No. 1.4e-99;  
 Matches 329; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63  
 |||  
 Db 1 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNDKEPPT 60  
 QY 64 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTEVEYISAHP 123  
 |||  
 Db 61 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRCPTVEYISSHP 120  
 QY 124 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 183  
 |||  
 Db 121 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 180

QY 184 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFA 243  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFT 240  
 QY 244 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQPKLIE 303  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQPKLIE 300  
 QY 304 FLSSFQKERTDDEQFADEKKNYLIKQIRDLLKKTAP 337  
 ||||||||||||||||||||||||||||  
 Db 301 FLSSFQKERTDDEQFADEKKNYLIKQIRDLLKKAAP 334

RESULT 3

MO25\_HUMAN

ID MO25\_HUMAN STANDARD; PRT; 341 AA.  
 AC Q9Y376;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE MO25 protein (CGI-66).  
 GN MO25.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20272150; PubMed=10810093;  
 RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;  
 RT "Identification of novel human genes evolutionarily conserved in  
 RT Caenorhabditis elegans by comparative proteomics.";  
 RL Genome Res. 10:703-713(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hypothalamus;  
 RA Jin W., Shi J., Ren S., Gu J., Fu S., Huang Q., Dong H., Yu Y., Fu G.,  
 RA Wang Y., Chen Z., Han Z.;  
 RT "A novel gene expressed in the human hypothalamus.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Duodenum;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,



RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- SIMILARITY: Belongs to the Mo25 family.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF151824; AAD34061.1; -.  
 DR EMBL; AF113536; AAF14873.1; -.  
 DR EMBL; BC020570; AAH20570.1; -.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 SQ SEQUENCE 341 AA; 39869 MW; EC710A528B6F9811 CRC64;

Query Match 81.0%; Score 1381; DB 1; Length 341;  
 Best Local Similarity 81.0%; Pred. No. 3.1e-81;  
 Matches 273; Conservative 31; Mismatches 29; Indels 4; Gaps 2;

Qy 4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59  
 || | ||||:|:| ||::|:| || | :|:| ||||:| ||||| |||||  
 Db 1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60  
 Qy 60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119  
 || ||||| ||||:| ||:| ||||| ||||| ||||| ||||:| |||||  
 Db 61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120  
 Qy 120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179  
 : ||||| ||:| || ||||| ||||| ||||| ||||:| || ||:|:| |||||  
 Db 121 CTQQNILFMLLKGYESPETALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180  
 Qy 180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239  
 ||||| ||||| ||:| ||:| || | :||| ||||| ||||| ||||:| |||||  
 Db 181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLDR 240  
 Qy 240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQP 299  
 ||| ||||| ||||| ||||| ||||| ||||| ||||:| ||||| ||||:| |||||  
 Db 241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVVFVANPNKTQPILDILLKNQA 300  
 Qy 300 KLIEFLSSSFQKERTDDEQFADEKKNYLIKQIRDLKKTAA 336  
 ||||| || :|:| || || ||:| ||||| :|  
 Db 301 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRDLKRPA 337

RESULT 4  
 MO25\_MOUSE  
 ID MO25\_MOUSE STANDARD; PRT; 341 AA.  
 AC Q06138;

DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE MO25 protein.  
 GN MO25 OR CAB39.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93119656; PubMed=8418809;  
 RA Miyamoto H., Matsushiro A., Nozaki M.;  
 RT "Molecular cloning of a novel mRNA sequence expressed in cleavage  
 RT stage mouse embryos.";  
 RL Mol. Reprod. Dev. 34:1-7(1993).  
 CC -!- FUNCTION: ONE OF THE FIRST GENES TO BE TRANSCRIBED DURING MOUSE  
 CC DEVELOPMENT, MAY PLAY SOME GENERAL FUNCTION.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -!- DEVELOPMENTAL STAGE: TRANSCRIBED DURING EARLY MOUSE DEVELOPMENT.  
 CC DETECTED AT ALL DEVELOPMENTAL STAGES FROM THE EGG THROUGH THE  
 CC BLASTOCYT, MOST ABUNDANT AT THE 2-CELL STAGE.  
 CC -!- SIMILARITY: Belongs to the Mo25 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; S51858; AAB24801.1; -.  
 DR MGD; MGI:107438; Cab39.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 SQ SEQUENCE 341 AA; 39842 MW; E7F668529D6FE811 CRC64;

Query Match 80.8%; Score 1376; DB 1; Length 341;  
 Best Local Similarity 80.7%; Pred. No. 6.5e-81;  
 Matches 272; Conservative 32; Mismatches 29; Indels 4; Gaps 2;

Qy 4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ--DKKTDKASEEVSKSLQAMKEILCGTNEK 59  
 || | |||:|:| ||:::| ||| :|:| |||:| ||||| |||||  
 Db 1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60  
  
 Qy 60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTEVEYI 119  
 || ||||| |||||:| ||| |:| ||||| ||||| ||||| |||||:| |||||  
 Db 61 EPQTEAVAQLAQELYNGLGLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTEVEYI 120  
  
 Qy 120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179  
 : ||||| |||||:| ||| ||||| ||||| ||||| |||||:| || |||:|:| |||||  
 Db 121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180  
  
 Qy 180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239  
 ||||| ||||| |||||:| |:| |||:| | :||| ||||| ||||| |||||:| |||  
 Db 181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLDR 240

Qy 240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299  
||| |||||||||||||||||||| | ||||||||||||| : : |||| : : |||||  
Db 241 HNFTIMTKYISKPENLKLMMNLLRDKSRIQFEAFHVFKVFVANPNKTQPIILDILLKNQT 300

Qy 300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336  
||| ||| | : ||| ||| ||| : ||| : |  
Db 301 KLIEFLSKFONDRTEDEFNDEKTYLVKOIRNLKRRAA 337

## RESULT 5

MO25 DROME

```

ID      MO25_DROME      STANDARD;      PRT;      339 AA.
AC      P91891; Q9VV85;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      MO25 protein (dMo25).
GN      MO25 OR CG4083.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96268479; PubMed=8672247;
RA      Nozaki M., Onishi Y., Togashi S., Miyamoto H.;
RT      "Molecular characterization of the Drosophila Mo25 gene, which is
RT      conserved among Drosophila, mouse, and yeast.";
RL      DNA Cell Biol. 15:505-509(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Berkeley;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA      Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA      Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -!- SIMILARITY: Belongs to the Mo25 family.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AB000402; BAA19098.1; -.  
 DR EMBL; AE003526; AAF49432.1; -.  
 DR FlyBase; FBgn0017572; Mo25.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 FT CONFLICT 51 51 Y -> H (IN REF. 1).  
 FT CONFLICT 102 102 V -> L (IN REF. 1).  
 SQ SEQUENCE 339 AA; 39385 MW; 5790BD91754C1C74 CRC64;

Query Match 65.2%; Score 1111; DB 1; Length 339;  
 Best Local Similarity 65.0%; Pred. No. 4.9e-64;  
 Matches 217; Conservative 59; Mismatches 54; Indels 4; Gaps 3;

Qy 4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63  
 |||| || |:| |:| ||: : || |:| :|| |:|:|:| :|| :| |:|:| ||  
 Db 1 MPLFGKSQKSPVELVKSLKEAINALEAGDRKVEKAQEDVSKNLVSIKNMLYGSSDAEPPA 60  
 Qy 64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122  
 : ||||:||||:| ||: || :| ||||| | ||||:|||||||  
 Db 61 DYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNVLRRQIGTRSPTVEYICTK 120  
 Qy 123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180  
 | ||| |: ||| |:||| | |||| |:| ||||:| |:| |:|:|:|  
 Db 121 PEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180  
 Qy 181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLQSENYVTKRQSLKLLGELILDR 239  
 ||||:||||:|||||:| |:|:| ||| | : |:| || ||||:| |||||:|  
 Db 181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLNSENYVTRRQSLKLLGELLDR 240  
 Qy 240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFASPHKTQPIVEILLKNQP 299  
 ||| :||:||||:|||||||:|:| ||||| |||||:|:| :||:| ||:|

Db 241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVVFVANPNKPKPILDILLRNQT 300  
 Qy 300 KLIEFLSSFQKERTDDEQFADEKKNYLIKQIRDLK 333  
 ||::||::| :|::||| ||| |||||::||  
 Db 301 KLVDFLTNTFHTDRSEDEQFNDEKAYLIKQIKELK 334

# RESULT 6

## MO2M\_CAEEL

ID MO2M\_CAEEL STANDARD; PRT; 338 AA.  
 AC 018211;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical MO25-like protein Y53C12A.4 in chromosome II.  
 GN Y53C12A.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Kershaw J., Lennard N.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: Belongs to the Mo25 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z99277; CAB16486.1; -.  
 DR PIR; T27129; T27129.  
 DR WormPep; Y53C12A.4; CE14890.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 338 AA; 39431 MW; 1D0C34A35D9116F5 CRC64;

Query Match 59.1%; Score 1006.5; DB 1; Length 338;  
 Best Local Similarity 57.2%; Pred. No. 2.2e-57;  
 Matches 191; Conservative 60; Mismatches 78; Indels 5; Gaps 1;

Qy 5 PLFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAMKEILCGTNEK 59  
 ||| | : | ||::|| | : | : : : : : | : | | : | | : | :  
 Db 4 PLFGKADKTPADVVKNLRLDALLVIDRHGTNTSERKVEKAIEETAKMLALAKTFIYGSDAN 63  
 Qy 60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119  
 || | | |||||:|: :| || | :|| |||| :|||:|||||:|||||:|  
 Db 64 EPNNEQVTQLAQEVYNANVLPMLIKHLHKFEFECKKDVASVFNNLLRRQIGTRSPTVEYL 123  
 Qy 120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179  
 :| | || || ||| | ||| || ||| :|| | :|:|:| | :| || :| :| ||

Db	124	AARPEILITLLLGYEQPDIALTCGSMRLREAVRHEHLARIVLYSEYFQRFFVFVQSDVFDI	183
Qy	180	ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENIVTKRQSLKLLGELILD	239
		:       :         :   :     :                       :	
Db	184	ATDAFSTFKDLMTKHKNMCAEYLDNNYDRFFGQYSALTNSENIVTRRQSLKLLGELL	243
Qy	240	HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFPASPHTQPIVEILLKNQP	299
		:         :                       :     :     :   :	
Db	244	HNFSTMNKYITSPENLKTVMELLRDKRNIQEAFHVKIFVANPNKPRPITDILTNRD	303
Qy	300	KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK	333
		:       :   :                                 :   :	
Db	304	KLVEFLTAFHNDRTNDEOFNDEKAYLIKOIOELR	337

YFV6 SCHPO

AC 09P708;

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Hypothetical protein C1834.06c in chromosome I.

OS Schizosaccharomyces pombe (Fission yeast).

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OX NCBI TaxID=4896;

RP SEQUENCE FROM N.A.

RX MEDLINE=21848401; PubMed=11859360;

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.

RA Oliver K., O'Neil S., Pearson D., Ouail M.A., Rabbinowitsch E.

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.

RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.

RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.

RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.

RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.

RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.

RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -!- SIMILARITY: Belongs to the Mo25 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL157734; CAB75774.1; -.  
 DR PIR; T50117; T50117.  
 DR GeneDB\_SPombe; SPAC1834.06c; -.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 329 AA; 38521 MW; 073DD0607A64C952 CRC64;

Query Match 49.0%; Score 834.5; DB 1; Length 329;  
 Best Local Similarity 51.5%; Pred. No. 1.9e-46;  
 Matches 169; Conservative 63; Mismatches 93; Indels 3; Gaps 2;

Qy 6 LFSKSHKNPAEIVKILKDNLAILE-KQDKKTDKASEEVSKSLQAMKEILCGTNEKEPTE 64  
 ||:| |: ::|: | || | || |: |||| | | :: |||| | | :  
 Db 4 LFNKRPKSTQDVVRCLCDNLPKLEINNDKK--KSFEEVSKCLQNLRVSLCGTAEVEPDAD 61

Qy 65 AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPH 124  
 |: |: ::| | | |: | ::|| || | |: :||| : :| |||:|: ||  
 Db 62 LVSDLFSQIYQSNLPFLLVRYLPKLEFESKKDTGLIFSALLRRHVASRYPTVDYMLAHPQ 121

Qy 125 ILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAF 184  
 | :|: | ::| | :||| || | ::| | | || :|:|:| ||||  
 Db 122 IFPVLVSYYRYQEVAFTAGSILRECSRHEALNEVLLNSRDFWTFFSLIQASSFDMASDAF 181

Qy 185 ATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFAI 244  
 :||| :| || |:|: ::| |: | |:| ||||| ||||| |||||:|:| | :  
 Db 182 STFKSILLNHKSQVAEFISYHFDEFFKQYTVLLKSENYVTKRQSLKLLGEILLNRANRSV 241

Qy 245 MTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVSPHKTPIVEILLKNQPKLIEF 304  
 ||:||| ||||| ||||| ||||| |||||:|:| |: :|:| |:| :| :  
 Db 242 MTRYISSAENLKLMMILLRDKSKNIQFEAFHVFKLFVANPEKSEEVIEILRNKSKLISY 301

Qy 305 LSSFQKERTDDEQFADEKKNYLIKQIRDL 332  
 ||:| :| :||| ||: ::||| |  
 Db 302 LSAFHTDRKNDEQFNDERAFVIKQIERL 329

# RESULT 8

DE76\_CHLPR

ID DE76\_CHLPR STANDARD; PRT; 321 AA.

AC Q9XFY6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Degreening related gene dee76 protein.

GN DEE76.  
 OS Chlorella protothecoides.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;  
 OC Chlorellaceae; Auxenochlorella.  
 OX NCBI\_TaxID=3075;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ACC25;  
 RX MEDLINE=20256472; PubMed=10798614;  
 RA Hortensteiner S., Chinner J., Matile P., Thomas H., Donnison I.S.;  
 RT "Chlorophyll breakdown in Chlorella protothecoides: characterization  
 RT of degreening and cloning of degreening-related genes.";  
 RL Plant Mol. Biol. 42:439-450(2000).  
 CC -!- SIMILARITY: Belongs to the Mo25 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ238632; CAB42595.1; -.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 SQ SEQUENCE 321 AA; 37262 MW; 918FD02964B09071 CRC64;

Query Match 45.5%; Score 776; DB 1; Length 321;  
 Best Local Similarity 52.0%; Pred. No. 9.8e-43;  
 Matches 156; Conservative 56; Mismatches 84; Indels 4; Gaps 3;

Qy	32	DKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDF	91
		:     :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	19	ESKQDRVVEDISKAIMSIKEAIFGEDEQSSSKEHAQGIASEACRVGLVSDLVTYLTVLDF	78
Qy	92	EGKKDVTQIFNNILRRQI--GTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLREC	149
		:           :   :   :   :   :   :   :   :   :   :   :   :	
Db	79	ETRKDVVQIFCAIIRITLEDGGR-PGRDYVLAHPDVLSTLFYGYEDPEIALNCGQMFREC	137
Qy	150	IRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTI	209
		:   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	138	IRHEDIAKFVLECNLFEELEKLNQSFVAVASDAFATFKDLLTRHKQLVAAFLQENYEDF	197
Qy	210	FEDYEKLLQSENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNI	269
		:      :      :      :      :      :      :      :      :	
Db	198	FSQLDKLLTSDNYVTRRQSLKLLGELLLDRVNVKIMMQYVSDVNNLILMMNLLKDSSRSI	257
Qy	270	QFEAFHVFKVFVSPHKTQPIVEILLKNQPKLIEFLSSFKERTDDEQFADEKKNYLIKQI	329
		:   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	258	QFEAFHVFKVFVANPNKTKPVADILVNNKNKLLTYLEDFHNDR-DDEQFKEEKAVIIEI	316

RESULT 9

MO2N\_ARATH

ID MO2N\_ARATH STANDARD; PRT; 343 AA.

AC Q9FGK3;



DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical MO25-like protein At5g47540.  
 GN AT5G47540 OR MNJ7.13.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Kaneko T., Kato T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: Belongs to the Mo25 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB025628; BAB09080.1; -.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 343 AA; 39457 MW; 46950D6A9A82FBB5 CRC64;

Query Match 42.7%; Score 728; DB 1; Length 343;  
 Best Local Similarity 43.2%; Pred. No. 1.2e-39;  
 Matches 147; Conservative 79; Mismatches 100; Indels 14; Gaps 4;

Qy 6 LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAMKEILCGTNE 58  
 || : ||::|: :| | : :| | | : |:|::: || || | :|  
 Db 4 LFKSKPRTPADLVRQTRDLLLFSRSTSLPDLRDSKRDEKMAELSRNIRDMKSILYGNSE 63  
  
 Qy 59 KEPPTAQAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTEVEY 118  
 || || ||| || : | || | : :| | ||: |: |:|: :| :|  
 Db 64 AEPVAEACAQLTQEFFKEDTLRLITCLPKLNLETRKDATQVVANLQRQVNSRLIASDY 123  
  
 Qy 119 ISAHPHILFMLLKGYEAPQIALRCGIMLRRECIRHEPLAKIILFSNQFRDFFKYVELSTFD 178  
 : |: : : :|::|: :|| | | |||||: :|| :| |: : || :| ||  
 Db 124 LEANIDLMDVLIIEGFENTDMALHYGAMFRECIHQIVAKYVLESDHVKKFFDYIQLPNFD 183  
  
 Qy 179 IASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDY-EKLLQSENYVTKRQSLKLLGELIL 237  
 ||:| | |||:| |||| | :|| :| | | || | ||:| ||:|:|:| |||:|  
 Db 184 IAADAAATFKELLTRHKSTVAEFLTKNEDWFFADYNSKLESSNYITRRQAIKLLGDILL 243  
  
 Qy 238 DRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKN 297  
 || | |:| |||:| :||::| ||||: | :|| |||||:| |: :| || ||:|  
 Db 244 DRSNSAVMTKYVSSRDNLRLILMNLLRESSKSIQIEAFHVFKLFAANQNKPADIVNILLVAN 303

Qy 298 QPKLIEFLSSFQKERTDDEQFADEKNYLIKQI-----RDL 332  
 : ||: |: : :: :||:| :|: ::::| |||  
 Db 304 RSKLLRLLADLKPKDK-EDERFEADKSQVLREIAALEPRDL 342

RESULT 10

MO2M\_ARATH

ID MO2M\_ARATH STANDARD; PRT; 343 AA.  
 AC Q9M0M4; O23570;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical MO25-like protein At4g17270.  
 GN AT4G17270 OR DL4670W.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=98121113; PubMed=9461215;  
 RA Bevan M., Bancroft I., Bent E., Love K., Goodman H.M., Dean C.,  
 RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,  
 RA Ridley P., Hudson S.-A., Patel K., Murphy G., Piffanelli P.,  
 RA Wedler H., Wedler E., Wambutt R., Weitzenegger T., Pohl T., Terryn N.,  
 RA Gielen J., Villarroel R., De Clercq R., van Montagu M., Lecharny A.,  
 RA Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S.,  
 RA Kotter P., Entian K.-D., Rieger M., Schaefer M., Funk B.,  
 RA Mueller-Auer S., Silvey M., James R., Monfort A., Pons A.,  
 RA Puigdomenech P., Douka A., Voukelatou E., Milioni D., Hatzopoulos P.,  
 RA Piravandi E., Obermaier B., Hilbert H., Duesterhoeft A., Moores T.,  
 RA Jones J.D.G., Eneva T., Palme K., Benes V., Rechmann S., Ansorge W.,  
 RA Cooke R., Berger C., Delseny M., Voet M., Volckaert G., Mewes H.-W.,  
 RA Klosterman S., Schueller C., Chalwatzis N.;  
 RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of  
 RT Arabidopsis thaliana.";  
 RL Nature 391:485-488(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
 RA Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
 RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
 RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,

RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,  
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,  
 RA Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,  
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,  
 RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,  
 RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,  
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
 RA Chefedor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,  
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,  
 RA Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;

RT "Sequence and analysis of chromosome 4 of the plant *Arabidopsis*  
 RT *thaliana*.";

RL Nature 402:769-777(1999).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;

RT "RIKEN *Arabidopsis* full length cDNA clones (RAFLs) sequenced by the  
 RT SSP consortium (Salk/Stanford/PGEC).";

RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.

CC -!- SIMILARITY: Belongs to the Mo25 family.

CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous  
 CC gene model prediction.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Z97343; CAB10508.1; ALT\_SEQ.

DR EMBL; AL161546; CAB78730.1; -.

DR EMBL; AF380659; AAK55740.1; -.

DR InterPro; IPR004892; Mo25.

DR Pfam; PF03204; Mo25; 1.

KW Hypothetical protein.

SQ SEQUENCE 343 AA; 39650 MW; D340B49A4924B7D1 CRC64;

Query Match 42.0%; Score 716.5; DB 1; Length 343;  
Best Local Similarity 42.9%; Pred. No. 6.5e-39;  
Matches 144; Conservative 78; Mismatches 105; Indels 9; Gaps 3;

```
Qy      6 LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAMKEILCGTNE 58
      || : ||:||: :| | : : :| : : :|:|||:: :| || | :|
Db      4 LFKSKPRTPADIVRQTRDLLLYADRSNSFPDLRESKREEKMVELSKSIRDCLKILYGNSE 63

Qy     59 KEPPTAVALAQELQYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEY 118
      || || ||| || : : | | : : | :|| ||: | :|: :| :|
Db     64 AEPVAEACAQLTQEFFKADTLRRLTSLPNLNLEARKDATQVVANLQRQQVNSRLIAADY 123

Qy    119 ISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFD 178
      : : : : | : | : || | | ||||| : || : | | : || | : |
Db    124 LESNIDLMDFLDVDFENTDMALHYGTMFRECIRHQIVAKYVLDSEHVKKFFYYIQLPNFD 183

Qy    179 IASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDY-EKLLQSENYVTKRQSLKLLGELIL 237
      ||:|| ||||:||||| ||:|| :| | | || |||:| ||:|:||:|||||::|
Db    184 IAADAAATFKELLTRHKSTVAEFLIKNEDWFFADYNSKLLSTNYITRRQAIKLLGDILL 243

Qy    238 DRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVSPHKTQPIVEILLKN 297
      || | |:||||:| :||:::|||||: | || |||||:|:||: :| | ||: |
Db    244 DRSNSAVMTKYVSSMDNLRILMNLRESSKTIQIEAFHVFKLFVANQNKPSDIANILVAN 303

Qy    298 QPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
      : ||: | : : : ||:| :| :::| :||
Db    304 RNKLLRLLADIKPDK-EDERFDADKAQVVREIANLK 338
```

#### RESULT 11

HYMA\_EMENI

ID HYMA\_EMENI STANDARD; PRT; 384 AA.

AC O60032;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Conidiophore development protein hymA.

GN HYMA.

OS Emericella nidulans (Aspergillus nidulans).

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiales; Trichocomaceae; Emericella.

OX NCBI\_TaxID=162425;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99126010; PubMed=9928930;

RA Karos M., Fischer R.;

RT "Molecular characterization of HymA, an evolutionarily highly conserved and highly expressed protein of Aspergillus nidulans.";

RL Mol. Gen. Genet. 260:510-521(1999).

CC -!- FUNCTION: Required for conidiophore development.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- SIMILARITY: Belongs to the Mo25 family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -





RESULT 13

HYM1\_YEAST

ID HYM1\_YEAST STANDARD; PRT; 399 AA.  
AC P32464;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE HYM1 protein.  
GN HYM1 OR YKL189W.  
OS *Saccharomyces cerevisiae* (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GRF88;  
RX MEDLINE=93348778; PubMed=8394042;  
RA Cheret G., Mattheakis L.C., Sor F.;  
RT "DNA sequence analysis of the YCN2 region of chromosome XI in  
RT *Saccharomyces cerevisiae*.";  
RL Yeast 9:661-667(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94205264; PubMed=8154185;  
RA Wiemann S., Voss H., Schwager C., Rupp T., Stegemann J.,  
RA Zimmermann J., Grothues D., Sensen C., Erfle H., Hewitt N.,  
RA Banrevi A., Ansorge W.;  
RT "Sequencing and analysis of 51.6 kilobases on the left arm of  
RT chromosome XI from *Saccharomyces cerevisiae* reveals 23 open reading  
RT frames including the FAS1 gene.";  
RL Yeast 9:1343-1348(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,  
RA Guerreiro P., Rodrigues-Pousada C.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.  
RN [4]  
RP GENE NAME.  
RX MEDLINE=20157038; PubMed=10655212;  
RA Dorland S., Deegenars M.L., Stillman D.J.;  
RT "Roles for the *Saccharomyces cerevisiae* SDS3, CBK1 and HYM1 genes in  
RT transcriptional repression by SIN3.";  
RL Genetics 154:573-586(2000).  
CC -!- SIMILARITY: Belongs to the Mo25 family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X69765; CAA49422.1; -.  
DR EMBL; X74151; CAA52249.1; -.  
DR EMBL; Z28189; CAA82032.1; -.

DR PIR; S34681; S34681.  
 DR SGD; S0001672; HYM1.  
 DR GO; GO:0005622; C:intracellular; IDA.  
 DR GO; GO:0016564; F:transcriptional repressor activity; IMP.  
 DR GO; GO:0007109; P:cytokinesis, completion of separation; IMP.  
 DR GO; GO:0008360; P:regulation of cell shape; IGI.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 SQ SEQUENCE 399 AA; 45853 MW; F48860754C892BA9 CRC64;

Query Match 28.5%; Score 485; DB 1; Length 399;  
 Best Local Similarity 33.0%; Pred. No. 4.3e-24;  
 Matches 113; Conservative 75; Mismatches 138; Indels 16; Gaps 6;

Qy 7 FSKSHKNPAEIVKILKDNLAILEK---QDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62  
 : | : | | : : : | | | | | : | | : | : |  
 Db 16 WKKNPKTPSDYARLIIEQLNKFSSPSLTQDNKR-KVQEECTKYLIGTKHFIVGDTDPHPT 74  
 Qy 63 TEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122  
 ||: :| : : : : | : :|| : : ||: | : ||:| : :  
 Db 75 PEAIDELYTAMHRADVFYELLHLHFVDLEFEARRECMLIFSICLGYSKDNKFVTVDYLVSQ 134  
 Qy 123 PHILFMLLKGYE-----APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELS 175  
 | : : :| | | | | : ||:| | :||| | ||: :|  
 Db 135 PKTISLMLRTAEVALQQKGCQDIFLTVGNMIECIKYEQLCRIILKDPQLWKFFFEFAKLG 194  
 Qy 176 TFDIASDAFATFKDLLTRHKVLVA-DFL--EQNYDTIFEDYEKLLQSENYVTKRQSLKLL 232  
 |:|: : : | | ||: :| | : ||: : ||||| |||  
 Db 195 NFEISTESLQILSAAFTAHPKLVSKEFFSNEINIIRFIKCINKLMAHGSYVTKRQSTKLL 254  
 Qy 233 GELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVSPHKTQPIVE 292  
 ||: | | |:| ||: |||||:| | : ||| |:| |||:|||| |:| |:| : :  
 Db 255 ASLIVIRSNNALMNIYINSPENLKLIMTMDKSKNLQLEAFNVFKVMVANPRKSKPVFD 314  
 Qy 293 ILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKK 334  
 ||:| |:| : : | : : | ||: : : :| | :  
 Db 315 ILVKNRDKLLTYFKTFGLD-SQDSTFLDEREFIVQEIDSLPR 355

RESULT 14  
 MO2L\_CAEEL  
 ID MO2L\_CAEEL STANDARD; PRT; 339 AA.  
 AC Q9T2M2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical MO25-like protein T27C10.3 in chromosome I.  
 GN T27C10.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Zhu H.J., Graves T., Hawkins M.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.



CC -!- SIMILARITY: Belongs to the Mo25 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF098504; AAC67411.1; -.  
 DR PIR; T33477; T33477.  
 DR WormPep; T27C10.3; CE19605.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 339 AA; 40232 MW; E7DA45CA33F2947E CRC64;

Query Match 8.4%; Score 143.5; DB 1; Length 339;  
 Best Local Similarity 19.3%; Pred. No. 0.02;  
 Matches 38; Conservative 50; Mismatches 76; Indels 33; Gaps 4;

Qy 159 ILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQ 218  
 :: :|:| | | | | : | : : : : | : : | : | :  
 Db 100 LMNTNKFRD-----FDVIQGTFTDLQIIFFTNHESANNFIKNNLPRFMQTLHKLIA 150  
 Qy 219 SENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFK 278  
 | : : : | | | | : | : : : : | : | : : : : : : : | : :  
 Db 151 CSNFFIQAKSFKFLNELFTAQTNYETRSLWMAEPAFIKLVVLAIQSNKHAVRSRAVSILE 210  
 Qy 279 VFVASPHKTQPIVEILLKNQPKLIEFL-----SSFQKERTDDEQFAD----- 320  
 :|: :| : : | : :|: || | : || | | :|  
 Db 211 IFIRNPRNSPEVHEFIGRNRNVLIAFFFNAPIHYQYQGSPNEKE---DAQYARMAYKLLN 267  
 Qy 321 ---EKNYLIKQIRDLKK 334  
 : : : | : | : :  
 Db 268 WDMQRPFTQEQLQDFEE 284

# RESULT 15

AKA9\_HUMAN

ID AKA9\_HUMAN STANDARD; PRT; 3911 AA.  
 AC Q99996; O14869; O43355; O94895; Q9UQH3; Q9UQQ4; Q9Y6B8; Q9Y6Y2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)  
 DE (PRKA9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor  
 DE protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)  
 DE (Hyperion protein) (Yotiao protein) (Centrosome- and Golgi-localized  
 DE PKN-associated protein) (CG-NAP).  
 GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 4).  
 RC TISSUE=Brain;  
 RX MEDLINE=98151389; PubMed=9482789;  
 RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;  
 RT "Yotiao, a novel protein of neuromuscular junction and brain that  
 RT interacts with specific splice variants of NMDA receptor subunit  
 RT NR1.";  
 RL J. Neurosci. 18:2017-2027(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.  
 RX MEDLINE=99219864; PubMed=10202149;  
 RA Witczak O., Skaalhegg B.S., Keryer G., Bornens M., Tasken K.,  
 RA Jahnsen T., Oerstavik S.;  
 RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring  
 RT protein located in the centrosome, AKAP450.";  
 RL EMBO J. 18:1858-1868(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RC TISSUE=Brain;  
 RX MEDLINE=99287934; PubMed=10358086;  
 RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;  
 RT "Characterization of a novel giant scaffolding protein, CG-NAP, that  
 RT anchors multiple signaling enzymes to centrosome and the Golgi  
 RT apparatus.";  
 RL J. Biol. Chem. 274:17267-17274(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Kemmner W.A., Deiss S., Schwarz U.;  
 RT "Cloning of Hyperion.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).  
 RC TISSUE=Gastric parietal cell;  
 RX MEDLINE=99115654; PubMed=9915845;  
 RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,  
 RA Trotter K.W., Milgram S.L., Goldenring J.R.;  
 RT "AKAP350, a multiply spliced protein kinase A-anchoring protein  
 RT associated with centrosomes.";  
 RL J. Biol. Chem. 274:3055-3066(1999).  
 RN [6]  
 RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).  
 RC TISSUE=Lymphoblast;  
 RA Hinds K., Sutterer C., Becker M., Hawkins M.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [7]  
 RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).  
 RC TISSUE=Lung;  
 RA Milgram S.L., Goldenring J.R., Schmidt P.H.;  
 RT "AKAP350: A multiply spliced family of proteins with centrosomal  
 RT association.";  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [8]  
 RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).  
 RC TISSUE=Brain;  
 RX MEDLINE=99087487; PubMed=9872452;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XI.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 5:277-286(1998).  
RN [9]  
RP SEQUENCE OF 17-1800 FROM N.A.  
RA Wu X., Graves T., Bradshaw H.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE  
CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND  
CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL  
CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN  
CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-  
CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR  
CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS  
CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.

CC -!- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N  
CC (PKN), PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)  
CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.

CC -!- SUBCELLULAR LOCATION: CENTROSOMAL IN MANY CELL TYPES AND  
CC CYTOPLASMIC IN PARIETAL CELLS.

CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=6;  
CC Name=1;  
CC IsoId=Q99996-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q99996-2; Sequence=VSP\_004102, VSP\_004107;  
CC Name=3; Synonyms=CG-NAP;  
CC IsoId=Q99996-3; Sequence=VSP\_004102, VSP\_004105, VSP\_004107;  
CC Name=4; Synonyms=Yotiao;  
CC IsoId=Q99996-4; Sequence=VSP\_004103, VSP\_004104;  
CC Name=5;  
CC IsoId=Q99996-5; Sequence=VSP\_004108;  
CC Name=6; Synonyms=AKAP350;  
CC IsoId=Q99996-6; Sequence=VSP\_004106, VSP\_004107, VSP\_004109;

CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY  
CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.

CC -!- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,  
CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A  
CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.

CC -!- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO  
CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.

CC -!- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR  
CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AJ131693; CAB40713.1; -.  
DR EMBL; AB019691; BAA78718.1; -.  
DR EMBL; AJ010770; CAA09361.1; -.  
DR EMBL; AF026245; AAB86384.1; -.

DR EMBL; AF083037; AAD22767.1; -.  
 DR EMBL; AC004013; AAB96867.1; ALT\_FRAME.  
 DR EMBL; AF091711; AAD39719.1; -.  
 DR EMBL; AB018346; BAA34523.1; -.  
 DR EMBL; AC000066; AAC60380.1; ALT\_FRAME.  
 DR Genew; HGNC:379; AKAP9.  
 DR MIM; 604001; -.  
 DR GO; GO:0005813; C:centrosome; TAS.  
 DR GO; GO:0005856; C:cytoskeleton; TAS.  
 DR GO; GO:0004973; F:N-methyl-D-aspartate receptor-associated pr. . .; TAS.  
 DR GO; GO:0005515; F:protein binding activity; TAS.  
 DR GO; GO:0007165; P:signal transduction; TAS.  
 DR GO; GO:0006832; P:small molecule transport; TAS.  
 DR GO; GO:0007268; P:synaptic transmission; TAS.  
 KW Coiled coil; Alternative splicing; Polymorphism.  
 FT DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING DOMAIN.  
 FT DOMAIN 164 914 COILED COIL (POTENTIAL).  
 FT DOMAIN 944 1022 COILED COIL (POTENTIAL).  
 FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).  
 FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).  
 FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).  
 FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).  
 FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).  
 FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).  
 FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).  
 FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).  
 FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).  
 FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).  
 FT DOMAIN 3587 3689 COILED COIL (POTENTIAL).  
 FT DOMAIN 3726 3730 POLY-LEU.  
 FT DOMAIN 203 292 GLN-RICH.  
 FT DOMAIN 321 1010 GLU-RICH.  
 FT DOMAIN 1846 2772 GLU-RICH.  
 FT VARSPLIC 17 28 Missing (in isoform 2 and isoform 3).  
 FT /FTId=VSP\_004102.  
 FT VARSPLIC 1637 1642 QLQEEI -> LATRRD (in isoform 4).  
 FT /FTId=VSP\_004103.  
 FT VARSPLIC 1643 3911 Missing (in isoform 4).  
 FT /FTId=VSP\_004104.  
 FT VARSPLIC 2175 2182 Missing (in isoform 3).  
 FT /FTId=VSP\_004105.  
 FT VARSPLIC 2175 2183 SADTFQKVE -> Q (in isoform 6).  
 FT /FTId=VSP\_004106.  
 FT VARSPLIC 2895 2907 VFGFYNMCFTLC -> GSSIPELAHSDAYQTREICSS  
 FT (in isoform 2, isoform 3 and isoform 6).  
 FT /FTId=VSP\_004107.  
 FT VARSPLIC 2895 2948 Missing (in isoform 5).  
 FT /FTId=VSP\_004108.  
 FT VARSPLIC 3901 3911 STTQFHAGMRR -> ALSLTTSWQHHSARPTAPLFFEILSH  
 FT SLG (in isoform 6).  
 FT /FTId=VSP\_004109.  
 FT VARIANT 1347 1347 K -> KQ.  
 FT /FTId=VAR\_010926.  
 FT CONFLICT 76 76 E -> Q (IN REF. 3).  
 FT CONFLICT 475 475 M -> I (IN REF. 3).  
 FT CONFLICT 554 554 E -> G (IN REF. 3).  
 FT CONFLICT 638 638 R -> S (IN REF. 3).

FT	CONFLICT	663	663	N -> S (IN REF. 3).
FT	CONFLICT	913	913	H -> N (IN REF. 3).
FT	CONFLICT	956	956	K -> N (IN REF. 3).
FT	CONFLICT	980	982	QKH -> PKP (IN REF. 1 AND 2).
FT	CONFLICT	997	997	Q -> P (IN REF. 1 AND 2).
FT	CONFLICT	1001	1001	Q -> P (IN REF. 1 AND 2).
FT	CONFLICT	1020	1020	N -> D (IN REF. 3).
FT	CONFLICT	1028	1028	V -> E (IN REF. 3).
FT	CONFLICT	1626	1626	R -> P (IN REF. 1 AND 2).
FT	CONFLICT	1703	1703	N -> T (IN REF. 3).
FT	CONFLICT	1707	1707	V -> G (IN REF. 3).
FT	CONFLICT	1802	1803	MISSING (IN REF. 5).
FT	CONFLICT	1843	1843	A -> P (IN REF. 3).

Query Match 7.5%; Score 128.5; DB 1; Length 3911;  
 Best Local Similarity 20.1%; Pred. No. 3.3;  
 Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;

Qy	18	VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG	77
		::            :   : : ::   :   :	
Db	664	IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE-----	710
Qy	78	LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI	125
		:: :      :: : ::       ::       : : :	
Db	711	--ISKLDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL	766
Qy	126	LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA	185
		:      :  :  :  :  :  :  :  :  :  :	
Db	767	LEKQMKKE-----NDLQEKFAQLEAEN-SILKDEKK	797
Qy	186	TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL	237
		: :    :  :  :  :  :  :  :  :  :  :  :	
Db	798	TLEDMLKIHTPVSQEERLIFLDSIKSKSDSVWEKEIEILIEENEDLKQQCIQLNEEIEK	857
Qy	238	DRHNFAIMTK-----YISKPENLKLMMNLLRD	264
		:  :        :  :  :  :  :  :  :	
Db	858	QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQLEYKSKLKALNEELHLQRI	917
Qy	265	KSPNIQFEA--FHVFKVVFVAPSKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD-	320
		:: :  :      :  :  :  :  :  :  :  :  :  :	
Db	918	NPTTVKMKSSVFEDEKTFVA---ETLEMGEVVEKDTTELMKLEVTKREKLELSQRLSDL	974
Qy	321	-----EKNYLIKQIRDLKK	334
		:  :  :  :  :	
Db	975	SEQLKQKHGEISFLNEEVKSLKQ	997

Search completed: January 7, 2004, 16:45:28  
 Job time : 20 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 16:44:17 ; Search time 41 Seconds  
(without alignments)  
2121.067 Million cell updates/sec

Title: US-10-088-872-2  
Perfect score: 1704  
Sequence: 1 MKKMPLFSKSHKNPAEIVKI.....FADEKNYLIKQIRDLKKTAP 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query					
No.	Score	Match	Length	DB	ID	Description	
-----							

1	1684	98.8	337	11	Q8BG52	Q8bg52 mus musculu
2	1669	97.9	334	11	Q91WB8	Q91wb8 mus musculu
3	1663	97.6	334	11	Q91YL0	Q91yl0 mus musculu
4	1462	85.8	289	4	Q96FG1	Q96fg1 homo sapien
5	1381	81.0	341	11	Q8VDZ8	Q8vdz8 mus musculu
6	1066.5	62.6	636	5	Q21643	Q21643 caenorhabdi
7	875	51.3	205	11	Q8K312	Q8k312 mus musculu
8	709.5	41.6	333	10	Q8H5L9	Q8h5l9 oryza sativ
9	671.5	39.4	345	10	Q8L9L9	Q8l9l9 arabidopsis
10	590	34.6	322	10	Q8LIF3	Q8lif3 oryza sativ
11	435	25.5	103	11	Q8K038	Q8k038 mus musculu
12	134.5	7.9	677	16	O25188	O25188 helicobacte
13	128	7.5	430	16	O26049	O26049 helicobacte
14	123.5	7.2	1285	16	Q9WXU3	Q9wxu3 thermotoga
15	120	7.0	1175	17	Q58914	Q58914 methanococc
16	119.5	7.0	1056	16	Q8REF7	Q8ref7 fusobacteri
17	119	7.0	1111	5	Q9VGE4	Q9vge4 drosophila
18	118.5	7.0	554	5	Q8IN90	Q8in90 drosophila
19	118.5	7.0	670	5	Q9VEC7	Q9vec7 drosophila
20	118.5	7.0	670	5	Q9NFM7	Q9nfm7 drosophila
21	117	6.9	808	5	Q8T133	Q8t133 dictyosteli
22	117	6.9	808	5	Q9GSH4	Q9gsh4 dictyosteli
23	116.5	6.8	1135	5	Q9NJQ4	Q9njq4 paramecium
24	116	6.8	911	16	Q8EUI7	Q8eui7 mycoplasma
25	116	6.8	1389	5	Q8I293	Q8i293 plasmodium
26	115.5	6.8	1111	5	Q9U0K5	Q9u0k5 plasmodium
27	115.5	6.8	1946	5	O97291	O97291 plasmodium
28	115	6.7	473	11	Q8R436	Q8r436 mus musculu
29	115	6.7	2518	5	Q8IEH2	Q8ieh2 plasmodium
30	114.5	6.7	1941	5	Q8IAK6	Q8iak6 plasmodium
31	114	6.7	743	13	Q9YGE7	Q9yge7 oncorhynchu
32	113.5	6.7	833	4	Q9UF54	Q9uf54 homo sapien
33	113.5	6.7	951	5	Q9VEC6	Q9vec6 drosophila
34	113.5	6.7	984	5	Q8IN89	Q8in89 drosophila
35	113	6.6	474	5	O97233	O97233 plasmodium
36	113	6.6	647	11	Q8CA10	Q8ca10 mus musculu
37	111.5	6.5	1925	5	Q8I2D1	Q8i2d1 plasmodium
38	111.5	6.5	2429	5	Q9VFB1	Q9vfb1 drosophila
39	111.5	6.5	2771	5	Q26216	Q26216 plasmodium
40	111	6.5	2166	16	O51465	O51465 borrelia bu
41	111	6.5	2819	16	Q98QP8	Q98qp8 mycoplasma
42	110	6.5	461	5	O77390	O77390 plasmodium
43	110	6.5	1183	2	O86064	O86064 helicobacte
44	110	6.5	1758	5	Q8I1K5	Q8ilk5 plasmodium
45	109.5	6.4	457	16	Q9PQM0	Q9pqm0 ureaplasma

#### ALIGNMENTS

##### RESULT 1

Q8BG52

ID Q8BG52 PRELIMINARY; PRT; 337 AA.

AC Q8BG52;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)





OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Salivary gland;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 DR EMBL; BC016128; AAH16128.1; -.  
 DR EMBL; AK076867; BAC36513.1; -.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 334 AA; 38718 MW; 822F04A87FB4EB6F CRC64;

Query Match 97.9%; Score 1669; DB 11; Length 334;  
 Best Local Similarity 98.5%; Pred. No. 1.3e-109;  
 Matches 329; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

Qy      4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
      |||
Db      1 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNDKEPPT 60

Qy     64 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHP 123
      |||
Db     61 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRCPTVEYISSHP 120

Qy    124 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 183
      |||
Db    121 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 180

Qy    184 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFA 243
      |||
Db    181 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFT 240

Qy    244 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFASPHKTQPIVEILLKNQPKLIE 303
      |||
Db    241 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFASPHKTQPIVEILLKNQPKLIE 300

Qy    304 FLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
      |||
Db    301 FLSSFQKERTDDEQFADEKNYLIKQIRDLKKAAP 334

```

RESULT 3

Q91YL0

ID Q91YL0 PRELIMINARY; PRT; 334 AA.

```
AC      Q91YL0;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE      Similar to hypothetical protein FLJ12577.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Strausberg R.;
RL      Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; BC016546; AAH16546.1; -.
DR      InterPro; IPR004892; Mo25.
DR      Pfam; PF03204; Mo25; 1.
KW      Hypothetical protein.
SQ      SEQUENCE      334 AA:  38761 MW:  5F9765360653750E CRC64:
```

Qy	4	MPLFSKSHKNPAEIVKILKDNLAIIEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT	63
		:	
Db	1	MPLFSKSHKNPAEIVKILKDNLAIIEKQDKKTDKASEEVSKSLQAMKEILCGTNDKEPPT	60
Qy	64	EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHP	123
		:	
Db	61	EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRCPTVEYISSHP	120
Qy	124	HILFMLLKGYEAPQIALRCGIMLRCEIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA	183
Db	121	HILFMLLKGYEAPQIALRCGIMLRCEIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA	180
Qy	184	FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLQSENYVTKRQSLKLLGELILDHRNFA	243
Db	181	FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLQSENYVTKRQSLKLRGELILDHRNFT	240
Qy	244	IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE	303
Db	241	IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE	300
Qy	304	FLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP	337
Db	301	FLSSFQKERTDDEQFADEKNYLIKQIRDLKKAAP	334

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC010993; AAH10993.1; -.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 289 AA; 33738 MW; F57B9EFCF6ABF2D7 CRC64;

Query Match 85.8%; Score 1462; DB 4; Length 289;  
 Best Local Similarity 99.7%; Pred. No. 3.8e-95;  
 Matches 288; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 49 MKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQ 108  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 MKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEEKKDVTQIFNNILRRQ 60  
  
 Qy 109 IGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDF 168  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 IGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDF 120  
  
 Qy 169 FKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQS 228  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 FKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQS 180  
  
 Qy 229 LKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVSPHKTQ 288  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 LKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVSPHKTQ 240  
  
 Qy 289 PIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 PIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 289

# RESULT 5

Q8VDZ8

ID Q8VDZ8 PRELIMINARY; PRT; 341 AA.  
 AC Q8VDZ8;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE MO25 protein.  
 GN CAB39.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC020041; AAH20041.1; -.

DR MGD; MGI:107438; Cab39.  
DR InterPro; IPR004892; Mo25.  
DR Pfam; PF03204; Mo25; 1.  
SQ SEQUENCE 341 AA; 39843 MW; E7FECA529D6FE811 CRC64;

Query Match 81.0%; Score 1381; DB 11; Length 341;  
Best Local Similarity 81.0%; Pred. No. 2.3e-89;  
Matches 273; Conservative 31; Mismatches 29; Indels 4; Gaps 2;

```
Qy      4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
      || | |||:|:| ||| |::|:| ||| ||| :||:|||||:| ||||| |||||
Db      1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy     60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
      || ||||| |||||:| ||| ||:||||| ||||| ||||| |||||:| |||||
Db     61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy    120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
      : ||||| ||||:|:| ||| ||||| ||||| |||||:| || |||:|:| |||||
Db    121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy    180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
      ||||| ||||| ||||:| |:| ||:| ||| :||| ||||| ||||| |||||:| |||
Db    181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLDDR 240

Qy    240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFVASPHKTQPIVEILLKNQP 299
      ||| ||||| ||||| ||||| ||||| ||||| |||||:|:| |||||:| |||||
Db    241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVVFVANPNKTQPILDILLKNQT 300

Qy    300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAA 336
      ||||| ||| :|:| ||| ||| ||:| |||||:| |
Db    301 KLIEFLSKFQNDRTEDQFNDEKTYLVKQIRDLKRAA 337
```

#### RESULT 6

Q21643

ID Q21643 PRELIMINARY; PRT; 636 AA.  
AC Q21643;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical 72.3 kDa protein.  
GN R02E12.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium."  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.





Best Local Similarity 44.0%; Pred. No. 3.7e-42;  
Matches 147; Conservative 67; Mismatches 109; Indels 11; Gaps 5;

```

Qy      6 LFSKSHKNPAEIVKILKDNLAILEKQ-----DKKTDKASEEVSKSLQAMKEILCGTNEK 59
      || : ||::|: :: | |: | :| :|::: :| || | |
Db      4 LFKSKPRTPADVVRQTRELLIFLDLHSGSRGGDAKREEKMAELSKNIRELKSILYGNES 63

Qy     60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
      || ||| || || : | || | :: | :|| ||: | :|: :: ||:
Db     64 EPVTEACVQLTQEFFRENTLRLLIICLPKLNLETRKDATQVVANLQRQQVSSKIVASEYL 123

Qy    120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
      |: :| |: || ||| | |||||::: | :| |: : || |:| |||
Db    124 EANKDLLDTLI-SYENMDIALHYGSMLECRIRHQSIA-YVLESDHMKKFFDYIQLPNFDI 181

Qy    180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYE-KLLQSENYVTKRQSLKLLGELILD 238
      |||| ||||:||||| ||:| | :|| | :: :|| | ||:||||:| ||::||
Db    182 ASDASATFKELLTRHKATVAEFLSKNYDWWFSEFNTRLLSSTNYITKRQAIKFLGDMLLD 241

Qy    239 RHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVVFASPHKTQPIVEILLKNQ 298
      | | :| :|:| :|| ::||| | ||| |||||:| |: :| :| ||: |:
Db    242 RSNSTVMMRYVSSKDNLMILMNLLRDSSKNIQIEAFHVFKLF AANKNKPTEVVNLTNR 301

Qy    299 PKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDL 332
      ||: | : |: :: |||| :| :||:| |
Db    302 SKLLRFFAGFKIDK--DEQFEADKEQVIKEISAL 333

```

# RESULT 9

Q8L9L9

```

ID   Q8L9L9      PRELIMINARY;          PRT;    345 AA.
AC   Q8L9L9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 0:0-0(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RT   "Full-Length cDNA from Arabidopsis thaliana.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY088359; AAM65898.1; -.
DR   InterPro; IPR004892; Mo25.
DR   Pfam; PF03204; Mo25; 1.

```

KW Hypothetical protein.

SQ SEQUENCE 345 AA; 39841 MW; 2C46A3D3DEBB47AA CRC64;

Query Match 39.4%; Score 671.5; DB 10; Length 345;

Best Local Similarity 42.9%; Pred. No. 1.8e-39;

Matches 140; Conservative 68; Mismatches 113; Indels 5; Gaps 2;

```
Qy      12 KNP AEIVKILKDNLAILEKQD----KKT DKASEEVSKSLQAMKEILCGTNEKEPPTEAVA 67
      | | | :| | :| :| | : : | :| | | | : : : | | | | :
Db      12 KTPQEVVKAIRDSL MALDTKT VVEVKALEKALEEVEKNFSSLRGILSGDGETEPNADQAV 71

Qy      68 QLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPHILF 127
      | | | | : : : | | : : : | :| : : : | :| | : |
Db      72 QLALEFCKEDVVS LVIHKLHILGWETR KDLLHCWSILLKQKVGDTYCCVQYFEEHFELLD 131

Qy     128 MLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATF 187
      | : | : :| | | | | | : | | | | | | | | :| | | | :|
Db     132 SLVV CYDNKEIALHCGSMLREC IKFPSLAKYILESACFELFFKFV ELPNFDVASDAFSTF 191

Qy     188 KDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFAIMTK 247
      | | | | : | : : : | : | : | | | | | | : : : | | |
Db     192 KDLLTKHDSVSVSEFLTSHYTEFFDVYERLLTSSNYVTRRQSLKLLSDFLLEPPNSHIMKK 251

Qy     248 YISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSS 307
      : | : | : : | | : | | | | | : | : | : | : | : |
Db     252 FILEVRYLKVIMTLLKDSSKNIQISAFHIFKIFVANPNKPQEVKII LARNHEKLELLELHD 311

Qy     308 FQKER-TDDEQFADEKKNYLIKQIRDL 332
      : : : | : | : | : | : | : |
Db     312 LSPGKGSEDDQFEEEEKELIIEEIQKL 337
```

#### RESULT 10

##### Q8LIF3

ID Q8LIF3 PRELIMINARY; PRT; 322 AA.  
AC Q8LIF3;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein (P0503D09.26 protein).  
GN OJ1316\_A04.9 OR P0503D09.26.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC  
RT clone:OJ1316\_A04.";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Katayose Y.;





DR EMBL; BC034159; AAH34159.1; -.  
DR InterPro; IPR004892; Mo25.  
DR Pfam; PF03204; Mo25; 1.  
SQ SEQUENCE 103 AA; 11291 MW; EA86A9F6E9E426E0 CRC64;

Qy            4 MPLFSKSHKNPAEIVKILKDNLAILLEKQDKTKASEEVSKSLQAMKEILCGTNEKEPPT 63  
| | | | | | | | | | | | | | | | | | | | : | | | |  
Db           1 MPLFSKSHKNPAEIVKILKDNLAILLEKQDKTKASEEVSKSLQAMKEILCGTNDKEPPT 60

Qy           64 EAVAQLAQELYSSGLLVTLIADLQLIDFEGK 94  
| | | | | | | | | | | | | | | | | |  
Db          61 EAVAQLAOELYSSGLLVTLIADLQLIDFEVK 91

## RESULT 12

DR SMART; SM00437; TOP1Ac; 1.  
 DR SMART; SM00436; TOP1Bc; 1.  
 DR SMART; SM00493; TOP1Rm; 1.  
 KW Hypothetical protein; Isomerase; Complete proteome.  
 SQ SEQUENCE 677 AA; 77677 MW; 4B285B81F1092BB4 CRC64;

Query Match 7.9%; Score 134.5; DB 16; Length 677;  
 Best Local Similarity 21.6%; Pred. No. 0.24;  
 Matches 88; Conservative 58; Mismatches 134; Indels 127; Gaps 16;

Qy 7 FSKSHKNPA-EIVKILKDNL-----AILEKQDKK---TDKASEEVSKSLQAMKE 51  
 | | | : : | | | | : : | | | : | | | :  
 Db 222 FKFCDKNEASQFLKDLKGLGSMVSLVSLKESLSNKEPKPKPFTTSKLLSQASKSLKI--- 278  
 Qy 52 ILCGTNEKEPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGT 111  
 | | : | | | | : | : | : | : | : | : | :  
 Db 279 -----PTKEIAQLAQKLFEAGLITYHRTDSEFLSPEYLKEHEVFFFEPIY----- 322  
 Qy 112 RSPTV----EYIS-----AHPHILFMLLKGYEAPQIALRCGIMLRECIRHE 153  
 | : | | : | | | | | : | : | : | : | :  
 Db 323 --PSVYQYREYKAGKNSQAEAEHAIRITHPHALKDLEKVCSDAKISEELALKLYQLIYTN 380  
 Qy 154 PL---AKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIF 210  
 : : : | : | : | | | : | : | : | : | :  
 Db 381 TICSQSRNALY-NQYDCIFK-----IKSESFKLSFKLLKEKGFLIEELIQGKEEIN 431  
 Qy 211 EDYEKLLQSENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQ 270  
 : | : | : | : | : : : : | : | : | :  
 Db 432 RE-EQSEIEIENFSLKENDSVPLKEVFIKK-----IEKPSPKPYKESAFIPLLESEG---- 481  
 Qy 271 FEAFHVFKVVFVASPHKTQPIVEILLKNQ-----PKLIEFLSSFQKERTDD- 315  
 : | : : | | : : : | : | : | : : | :  
 Db 482 -----IGRPSTYASFLDLLLKRKYISIDTKTNAITPTSQGLEVISFFKKDKKEVDF 531  
 Qy 316 -----EQF-----ADEKNYLIKQIRDLKKTA 336  
 : | : | : : : | : | : | :  
 Db 532 IALTSKDKSKLGNTTKQFEECLDLIMRGEASYEKFMLEVISKLKSTA 578

# RESULT 13

O26049

ID O26049 PRELIMINARY; PRT; 430 AA.  
 AC O26049;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein HP1520.  
 GN HP1520.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,

RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathley L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*  
 RT *pylori*";  
 RL Nature 388:539-547(1997).  
 DR EMBL; AE000650; AAD08565.1; -.  
 DR TIGR; HP1520; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 430 AA; 50573 MW; 23DC6FE5E956B629 CRC64;

Query Match 7.5%; Score 128; DB 16; Length 430;  
 Best Local Similarity 20.9%; Pred. No. 0.39;  
 Matches 82; Conservative 73; Mismatches 135; Indels 102; Gaps 20;

Qy 7 FSKSHKNPAEI----VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62  
 | : | : || ||| : | : | : : | : | : : | :  
 Db 60 FYPNRKSKIEIEFNNGEKILKENVAVFHSYDE--EFSSSDSVTTTFMAKSDL-----KQQY 111  
  
 Qy 63 TEAVAQLAQELYSSGLLVTL--IA-----DLQLIDFEGKKDVTQIFNNILR----- 106  
 : : | : | || : | || : : | | | : | : | |  
 Db 112 DNILLELEKE--KKALLKSLRDIASGFDYEBEIKTIKNEKNKSFYEILDNHLTEIESSEK 169  
  
 Qy 107 -----RQIGTRSP TV-EYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKII 159  
 | | | | : : : | : : | : | : | : | : | :  
 Db 170 HYSFKYRDI FDGSKKVKDFVNKHHDLIEQYFNKYQ-----ELLSQSK 211  
  
 Qy 160 LF-----SNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQ----- 204  
 : | : | | | : | : | : | : : : : | : | : |  
 Db 212 IFKHMNSGDFGTNHADDLKKALENNRFFKANHSLKIAGEEITNYQKL-SDIFENEKNRIL 270  
  
 Qy 205 NYDTIFEDYEKLLQSENVYTKRQSLKLLGELI-----LDRHNF--AIMTKYISKP 252  
 | : : | : : | : | : : | : | : | : | : | : :  
 Db 271 NNEELKESFDKI---EKVINANKELKAFKDAISKDNTLLTEFLDYDSFRKKVLF SYLKQV 327  
  
 Qy 253 -ENLKLMMNLLRDKSPNIQFEAFHVFKV FVASPHKTQPIVEILLKNQPKLIEFLSSFQKE 311  
 : | : | : : | | : | : | : : : : | | | : | :  
 Db 328 IQNVKSLVNL YREKKPEIE---EIIKQASKDQKEWESVIEIF--NQRFLVPFKVELQNQ 381  
  
 Qy 312 R-----TDDEQ----FADEKNYLIKQIRDLKK 334  
 : : | | | : | : | : | : | : | :  
 Db 382 KDILLNKDAAQFRFIFSDDNQDMNVQKEDLQK 413

# RESULT 14

Q9WXU3

ID Q9WXU3 PRELIMINARY; PRT; 1285 AA.

AC Q9WXU3;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE COME protein, putative.



## 058914

Query Match 7.0%; Score 120; DB 17; Length 1175;  
Best Local Similarity 21.5%; Pred. No. 4.8;  
Matches 76; Conservative 58; Mismatches 131; Indels 88; Gaps 15;

Qy	7	FSKSHKNPAEIVKILKD-NLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT	65
		:  : :      :      :  :   : ::  :	
Db	232	FNKFREEHQDFDKYLTDENIAFRPHVMKKFDEFAENIKKVIAELE---GSKYKYPGLPG	287
Qy	66	VAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTEVEYISAHPHI	125
		: ::    :  ::  :: :   : :	
Db	288	V-----LYFLGMEDAYSRYIELWKNEGEKGEEKLYNALI-ESLENRKENLEF-----	333
Qy	126	LFMLLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFK----YVELSTFDIA-	180
		: :    :              :	
Db	334	-----GITKKVIDKFIAQKEEFREFLNKYAVVYELSFAFKLEK	370
Qy	181	-----SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSL-----	229
		:: :      :  : : ::   : :	
Db	371	IKEQYEKEFINLDNIIKNPYILVED-LKEN-----DSFERIIFEELDSWERRRLGDKFNP	424

Qy 230 -----KLLGELILDRH---NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAF---- 274  
 || | || | | | | | | : || : | | : |  
 Db 425 YSPYRV RALLVE-ILKRHLSSGNTTISTK-----DLKDFFEKMDKDIVKITFDEFLRII 477  
 Qy 275 HVFKV FVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIK 327  
 : | : : | : : : : | : | | | : : : | : | : ||| : |  
 Db 478 EEYKDIIS--EKVEIVKKEVKNNENKEIIEFTLKEIREYEEIIENTINYLLK 528

Search completed: January 7, 2004, 16:48:05  
 Job time : 56 secs